(19) DANMARK (10) DK 2017 70434 A1

(12) PATENTANSØGNING

Patent- og
Varemærkestyrelsen

(51) Int.Cl.: *G10L 15/20 (2006.01)*

(21) Ansøgningsnummer: **PA 2017 70434**

(22) Indleveringsdato: **2017-06-02**

(24) Løbedag: **2017-06-02**

(41) Alm. tilgængelig: **2018-11-17**

(43) Publiceringsdato: **2018-12-04**

(30) Prioritet:
2017-05-16 US 62/507,151

(71) Ansøger:
Apple Inc., 1 Infinite Loop Cupertino 95014 CA California, USA

(72) Opfinder:
Thomas R. Gruber, c/o Apple Inc. 1 Infinite Loop Cupertino 95014 CA California, USA
Yoon Kim, c/o Apple Inc. 1 Infinite Loop Cupertino 95014 CA California, USA
David A. Carson, c/o Apple Inc. 1 Infinite Loop Cupertino 95014 CA California, USA
Charles Srisuwananukorn, c/o Apple Inc. 1 Infinite Loop Cupertino 95014 CA California, USA
Justin G. Binder, c/o Apple Inc. One Apple Park Way Cupertino 95014 CA California, USA

(74) Fuldmægtig:
COPA COPENHAGEN PATENTS K/S, Rosenørns Allé 1, 2. sal, 1970 Frederiksberg C, Danmark

(54) Titel: **FAR-FIELD EXTENSION FOR DIGITAL ASSISTANT SERVICES**

(56) Fremdragne publikationer:
US 2015/230095 A1
WO 2016111881 A1
US 2016/104486 A1
US 2015/373183 A1
EP 2339576 A2
US 7606712 B1
US 2016/072940 A1

(57) Sammendrag:
Systems and processes for operating an intelligent automated assistant to provide extension of digital assistant services are provided. An example method includes, at an electronic device having one or more processors, receiving, from a first user, a first speech input representing a user request. The method further includes obtaining an identity of the first user; and in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device. The method further includes receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide the response to the first electronic device, the response to the user request from the second electronic device or the third electronic device. The method further includes providing a representation of the response to the first user.

Fortsættes...

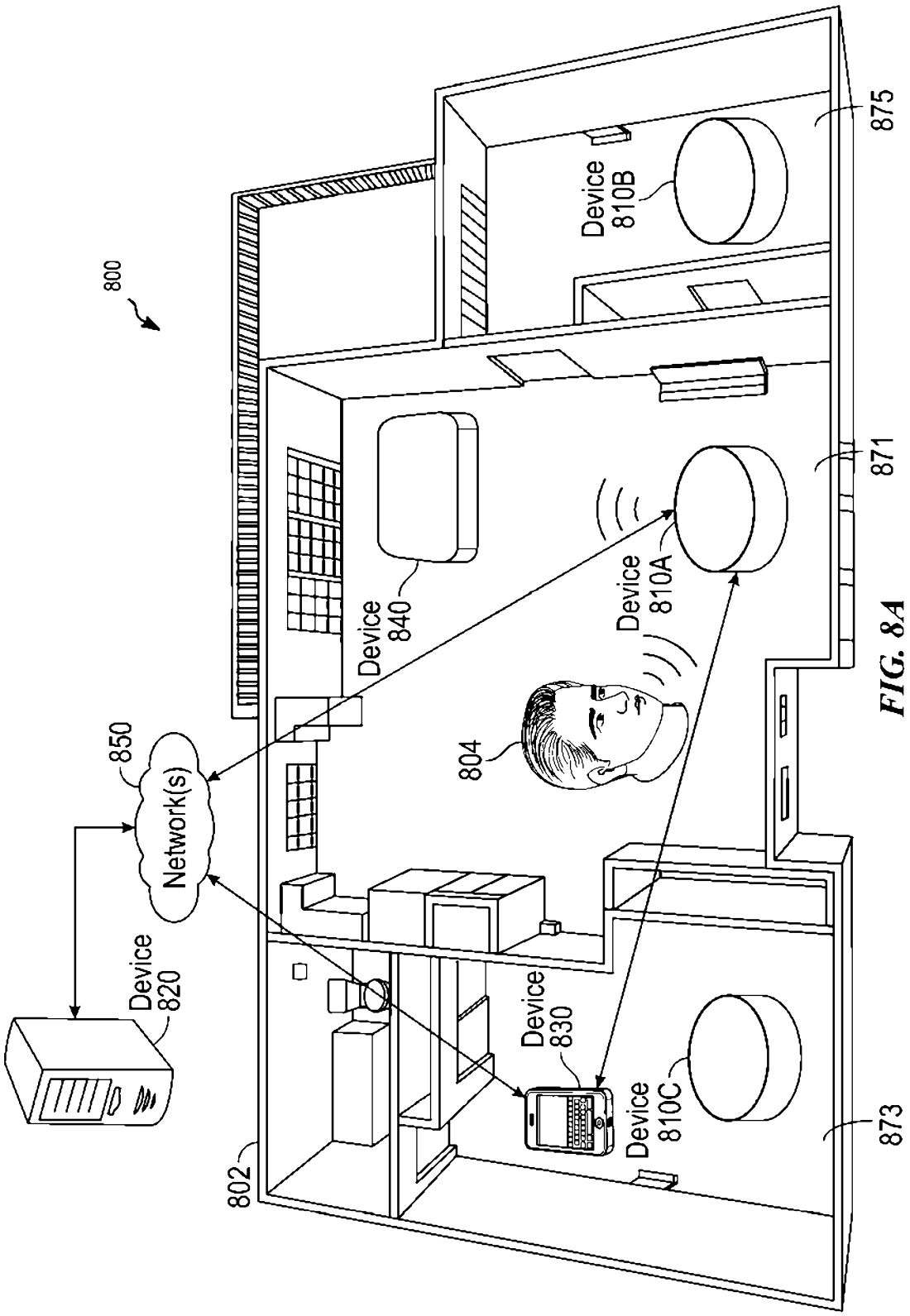

800
875
Device
810B
871
Device
840
Device
810A
804
850
Network(s)
Device
820
Device
830
Device
810C
802
873

FIG. 8A

FAR-FIELD EXTENSION FOR DIGITAL ASSISTANT SERVICES

[0001] This application claims priority to U.S. provisional patent application 62/507,151, entitled "FAR-FIELD EXTENSION FOR DIGITAL ASSISTANT SERVICES", filed on May 16, 2017, the content of which is hereby incorporated by reference in their entirety.

Field

[0002] This relates generally to intelligent automated assistants and, more specifically, to far field extension for digital assistant services.

Background

[0003] Intelligent automated assistants (or digital assistants) can provide a beneficial interface between human users and electronic devices. Such assistants can allow users to interact with devices or systems using natural language in spoken and/or text forms. For example, a user can provide a speech input containing a user request to a digital assistant operating on an electronic device. The digital assistant can interpret the user's intent from the speech input and operationalize the user's intent into tasks. The tasks can then be performed by executing one or more services of the electronic device, and a relevant output responsive to the user request can be returned to the user.

[0004] Using a digital assistant typically require direct interaction between the user and the digital assistant. For example, the user may be required to be in close proximity (e.g., in the same room) with the electronic device on which the digital assistant operates. The digital assistant may thus directly receive the user's speech input via its microphone and provide response to the user via its speaker. Under certain circumstances, requiring the user to be in close proximity with the electronic device may cause difficulty and inconvenience for the user to interact with the digital assistant. For example, if the user and the electronic device on which the digital assistant operates are separated beyond a distance (e.g., in different rooms) such that the digital assistant is incapable of, or has difficult of, receiving the user's speech input, the digital

assistant may be incapable of providing digital assistant services to the user. Thus, techniques for far-field extension of digital assistant services are desired.

[0005] Furthermore, different types of electronic devices may have different capabilities. As a result, digital assistant services provided at different devices may be different. Certain digital assistant services may not be provided at certain devices due to device capability limitations. For example, while a digital assistant operating on a smartphone device may output voice reading of text messages, a digital assistant operating on a TV set-top box may be incapable of doing the same due to device limitations. Thus, it is desired to provide digital assistant services using multiple devices to mitigate the device capability limitation.

Summary

[0006] Systems and processes for providing digital assistant services are provided.

[0007] Example methods are disclosed herein. An example method includes, at an electronic device having one or more processors, receiving, from a first user, a first speech input representing a user request. The method further includes obtaining an identity of the first user; and in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device. The method further includes receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide the response to the first electronic device, the response to the user request from the second electronic device or the third electronic device. The method further includes providing a representation the response to the first user.

[0008] Example non-transitory computer-readable media are disclosed herein. An example non-transitory computer-readable storage medium stores one or more programs. The one or more programs comprise instructions, which when executed by one or more processors of an electronic device, cause the electronic device to receive, from a first user, a first speech input representing a user request. The one or more programs further comprise instructions that cause the electronic device to obtain an identity of the first user; and in accordance with the user identity, provide a representation of the user request to at least one of a second electronic device or a third electronic device. The one or more programs further comprise instructions that cause

the electronic device to receive, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide the response to the first electronic device, the response to the user request from the second electronic device or the third electronic device. The one or more programs further comprise instructions that cause the electronic device to provide a representation of the response to the first user.

[0009] Example electronic devices are disclosed herein. An example electronic device comprises one or more processors; a memory; and one or more programs, where the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs including instructions for receiving, from a first user, a first speech input representing a user request. The one or more programs further include instructions for obtaining an identity of the first user; and in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device. The one or more programs further include instructions for receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide the response to the first electronic device, the response to the user request from the second electronic device or the third electronic device. The one or more programs further include instructions for providing a representation of the response to the first user.

[0010] An example electronic device comprises means for receiving, from a first user, a first speech input representing a user request. The electronic device further includes means obtaining an identity of the first user; and in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device. The electronic device further includes means for receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide the response to the first electronic device, the response to the user request from the second electronic device or the third electronic device. The electronic device further includes means for providing a representation of the response to the first user.

[0011] Example methods are disclosed herein. An example method includes, at an electronic device having one or more processors, receiving a notification of an event associated with a first user. The method further includes, in response to receiving the notification, outputting an

indication of the notification. The method further includes receiving one or more speech inputs; and in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device. The method further includes, in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

[0012] Example non-transitory computer-readable media are disclosed herein. An example non-transitory computer-readable storage medium stores one or more programs. The one or more programs comprise instructions, which when executed by one or more processors of an electronic device, cause the electronic device to receive a notification of an event associated with a first user. The one or more programs further include instructions that cause the electronic device to output an indication of the notification in response to receiving the notification. The one or more programs further include instructions that cause the electronic device to receive one or more speech inputs; and in accordance with the one or more speech inputs, determine whether the notification is to be provided at the first electronic device. The one or more programs further include instructions that cause the electronic device to, in accordance with a determination that the notification is to be provided at the first electronic device, provide the notification at the first electronic device.

[0013] Example electronic devices are disclosed herein. An example electronic device comprises one or more processors; a memory; and one or more programs, where the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs including instructions for receiving a notification of an event associated with a first user. The one or more programs including instructions for, in response to receiving the notification, outputting an indication of the notification. The one or more programs including instructions for receiving one or more speech inputs; and in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device. The one or more programs including instructions for, in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

[0014] An example electronic device comprises means for receiving a notification of an

event associated with a first user. The electronic device further includes means for, in response to receiving the notification, outputting an indication of the notification. The electronic device further includes means for receiving one or more speech inputs; and in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device. The electronic device further includes means for, in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

[0015] Example methods are disclosed herein. An example method includes, at an electronic device having one or more processors, receiving, from a first user, a first speech input representing a user request. The method further includes obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device. The method further includes, in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request. The method further includes causing the second electronic device to provide at least a portion of the response to the first user.

[0016] Example non-transitory computer-readable media are disclosed herein. An example non-transitory computer-readable storage medium stores one or more programs. The one or more programs comprise instructions, which when executed by one or more processors of an electronic device, cause the electronic device to receive, from a first user, a first speech input representing a user request. The one or more programs further include instructions that cause the electronic device to obtain capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device. The one or more programs further include instructions that cause the electronic device to, in accordance with the capability data, identify, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request. The one or more programs further include instructions that cause the electronic device to provide at least a portion of the response to the first user.

[0017] Example electronic devices are disclosed herein. An example electronic device comprises one or more processors; a memory; and one or more programs, where the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs including instructions for receiving, from a first user, a first speech input representing a user request. The one or more programs further include instructions for obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device. The one or more programs further include instructions for, in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request. The one or more programs further include instructions for causing the second electronic device to provide at least a portion of the response to the first user.

[0018] An example electronic device comprises means for receiving, from a first user, a first speech input representing a user request. The electronic device further includes means for obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device. The electronic device further includes means for, in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request. The electronic device further includes means for causing the second electronic device to provide at least a portion of the response to the first user.

[0019] Techniques for far-field extension of digital assistant services by one or more service-extension devices can improve the user-interaction interface. For example, using one or more service-extension devices, a user is no longer required to be in close proximity (e.g., in the same room) with an electronic device for receiving digital assistant services provided by the digital assistant operating on the electronic device. Further, the service-extension devices can flexibly obtain responses to user requests from a device disposed in the vicinity of the user and/or a device disposed remotely, depending on the content of the user request. For example, if the user requests personal information (e.g., calendar events), a service-extension device may obtain a response from a device disposed in the vicinity of the user (e.g., the user's smartphone), rather

than a remote device, thereby reducing the time required for providing services to the user. Under some circumstances, obtaining a response from a local device may also alleviate privacy concerns because sensitive or confidential information may be contained in a communicated between local devices. Further, the ability to obtain responses from different devices enhances the capability of a service-extension device to provide responses to a user. For example, if user-requested information cannot be obtained from one device (e.g., the user's smartphone), the service-extension device may obtain the response from another device (e.g., a server). As a result, a service-extension device can dynamically obtain responses from one or more devices, and efficiently extend digital assistant services from multiple devices.

[0020] One or more service-extension devices can further extend digital assistant services to enhance the continuity in providing digital assistant services. For example, one or more service-extension devices can determine whether a response to the user request (e.g., playing music) is to be provided at any particular service-extension device or at another electronic device, depending on the user's location, movement, preferences, etc. This capability of selecting a best device to provide service extension enhances the continuity for providing digital assistant services among multiple devices and further improves the user-interaction interface. Moreover, one or more service-extension devices can be shared by multiple users (e.g., family members) and the operation of the devices can be based on authentication of the multiple users. As a result, the same service-extension devices can extend digital assistant services from multiple electronic devices associated with multiple users. This capability of sharing service-extension devices enhances the efficiency in providing the digital assistant extension services.

[0021] Furthermore, techniques for providing a notification to the user using one or more service-extension devices can provide prompt notifications to the user in an extended distance. For example, a user may be separated from a user device for a distance and may thus be incapable of directly receiving notifications provided by the user device. One or more service-extension devices can receive notifications from the user device (e.g., the user's smartphone), and provide an audio and/or visual output associated with the notification to the user. Thus, the service-extension devices effectively extended the distance that a user device can provide notifications to the user.

[0022] Furthermore, techniques for providing digital assistant services using multiple devices can mitigate the device capability limitation. For example, a user device may not be able to provide services in response to user requests due to the limitation of its capability (e.g., small screen size, lack of requested information, etc.). The user device can identify another device that is capable of providing the services and cause the other device to provide the requested services to the user. The ability to identify another device that is capable of providing the requested services leverages the capabilities of a collection of devices to provide digital assistant services to the user, and enhances the user-interaction efficiency by reducing the user's burden to seek for a suitable device.

[0023] Furthermore, these techniques enhance the operability of the device and makes the user-device interface more efficient, which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

Brief Description of the Drawings

[0024] FIG. 1 is a block diagram illustrating a system and environment for implementing a digital assistant, according to various examples.

[0025] FIG. 2A is a block diagram illustrating a portable multifunction device implementing the client-side portion of a digital assistant, according to various examples.

[0026] FIG. 2B is a block diagram illustrating exemplary components for event handling, according to various examples.

[0027] FIG. 3 illustrates a portable multifunction device implementing the client-side portion of a digital assistant, according to various examples.

[0028] FIG. 4 is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface, according to various examples.

[0029] FIG. 5A illustrates an exemplary user interface for a menu of applications on a portable multifunction device, according to various examples.

[0030] FIG. 5B illustrates an exemplary user interface for a multifunction device with a touch-sensitive surface that is separate from the display, according to various examples.

[0031] FIG. 6A illustrates a personal electronic device, according to various examples.

[0032] FIG. 6B is a block diagram illustrating a personal electronic device, according to various examples.

[0033] FIG. 7A is a block diagram illustrating a digital assistant system or a server portion thereof, according to various examples.

[0034] FIG. 7B illustrates the functions of the digital assistant shown in FIG. 7A, according to various examples.

[0035] FIG. 7C illustrates a portion of an ontology, according to various examples.

[0036] FIGs. 8A-8B illustrate functionalities of providing digital assistant services at a first electronic device based on a user input, according to various examples.

[0037] FIGs. 9A-9C illustrate functionalities of obtaining an identity of a user at a first electronic device, according to various examples.

[0038] FIGs. 10A-10C illustrate functionalities of providing digital assistant services based on a user request for information, according to various examples.

[0039] FIGs. 11A-11D illustrate functionalities of providing digital assistant services based on a user request for performing a task, according to various examples.

[0040] FIGs. 12A-12C illustrate functionalities of providing digital assistant services based on a user request for information, according to various examples.

[0041] FIGs. 13A-13B illustrate functionalities of providing digital assistant services at a first electronic device or additional electronic devices, according to various examples.

[0042] FIG. 14 illustrates functionalities of providing continuity of digital assistant services between different electronic devices, according to various examples.

[0043] FIGs. 15A-15G illustrate functionalities of providing digital assistant services based on a notification of an event, according to various examples.

[0044] FIGs. 16A-16I illustrate a process for providing a digital assistant service at a first electronic device based on a user input, according to various examples.

[0045] FIGs. 17A-17D illustrate a process for providing digital assistant services based on a notification of an event, according to various examples.

[0046] FIGs. 18A-18E illustrate functionalities for providing digital assistant services based on capabilities of multiple electronic devices, according to various examples.

[0047] FIGs. 19A-19D illustrate a process for providing digital assistant services based on capabilities of multiple electronic devices, according to various examples.

Detailed Description

[0048] In the following description of examples, reference is made to the accompanying drawings in which are shown by way of illustration specific examples that can be practiced. It is to be understood that other examples can be used and structural changes can be made without departing from the scope of the various examples.

[0049] The present disclosure provides techniques for far-field extension of digital assistant services by one or more service-extension devices. As described, using service-extension devices can improve the user-interaction interface. In some examples, a first electronic device can be a service-extension device. The first electronic device can receive a speech input representing a user request. The first electronic device can obtain an identity of the user based on, for example, authentication of the user by a second electronic device and/or a third electronic device. In some examples, the second electronic device can be a device disposed remotely from the first electronic device (e.g., a remote server); and the third electronic device can be a device disposed in the vicinity of the first electronic device (e.g., the user's smartphone). After the identified is obtained, the first electronic device can provide a representation of the user request

to at least one of the second electronic device and the third electronic device. One or both of the second electronic device and the third electronic device can determine whether to provide a response to the first electronic device. The first electronic device (e.g., a service-extension device) can receive the response and provide a representation of the response to the user. As such, the first electronic device effectively extends the digital assistant services provided by one or both of the second electronic device and the third electronic device.

[0050] The present disclosure further provides techniques for providing notifications using one or more service-extension devices. As described above, using one or more service-extension devices, notifications can be provided to the user promptly in an extended distance. In some examples, a first electronic device can receive a notification from another device (e.g., a user's smartphone) and output an indication (e.g., a beep) of the notification. The first electronic device may receive one or more speech inputs inquiring about the indication and instructing the first electronic device to perform an operation of the notification (e.g., outputting the notification). The first electronic device can determine whether the notification should be provided; and provide the notification according to the determination.

[0051] The present disclosure further provides techniques for providing digital assistant services using multiple devices. As described above, providing digital assistant services using multiple devices can mitigate the device capability limitation. In some examples, a first electronic device receives a speech input representing a user request and obtains capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device. The capability data can include device capabilities and informational capabilities. In accordance with the capability data, the first electronic device can identify a second electronic device for providing at least a portion of a response to the user request; and cause the second electronic device to provide at least a portion of the response.

[0052] 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 are both inputs

and, in some cases, are separate and different inputs.

[0053] 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.

[0054] 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

[0055] FIG. 1 illustrates a block diagram of system 100 according to various examples. In some examples, system 100 implements a digital assistant. The terms "digital assistant," "virtual assistant," "intelligent automated assistant," or "automatic digital assistant" 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 performs 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.

[0056] Specifically, a digital assistant is 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 seeks either an informational answer or performance of a task by the digital assistant. A satisfactory response to the user request includes a provision of the requested informational answer, a performance of the requested task, or a combination of the two. For example, a user asks the digital assistant a question, such as "Where am I right now?" Based on the user's current location, the digital assistant answers, "You are in Central Park near the west gate." The user also requests 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 sometimes interacts 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 also provides responses in other visual or audio forms, e.g., as text, alerts, music, videos, animations, etc.

[0057] As shown in FIG. 1, in some examples, a digital assistant is implemented according to a client-server model. The digital assistant includes 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 communicates with DA server 106 through one or more networks 110. DA client 102 provides client-side functionalities such as user-facing input and output processing and communication with DA server 106. DA server 106 provides server-side functionalities for any number of DA clients 102 each residing on a respective user device 104.

[0058] In some examples, DA server 106 includes 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 facilitates the client-facing input and output processing for DA server 106. One or more processing modules 114 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 communicates with external services 120 through network(s) 110 for task completion or information acquisition. I/O interface to external services 118 facilitates such communications.

[0059] User device 104 can be any suitable electronic device. In some examples, user device is 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 is, for example, a mobile telephone that also contains other functions, such as PDA and/or music player functions. Specific examples of portable multifunction devices include the iPhone®, iPod Touch®, and iPad® devices from Apple Inc. of Cupertino, California. Other examples of portable multifunction devices include, without limitation, laptop or tablet computers. Further, in some examples, user device 104 is a non-portable multifunctional device. In particular, user device 104 is a desktop computer, a game console, a television, or a television set-top box. In some examples, user device 104 includes a touch-sensitive surface (e.g., touch screen displays and/or touchpads). Further, user device 104 optionally includes 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.

[0060] Examples of communication network(s) 110 include local area networks (LAN) and wide area networks (WAN), e.g., the Internet. Communication network(s) 110 is 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.

[0061] Server system 108 is implemented on one or more standalone data processing apparatus or a distributed network of computers. In some examples, server system 108 also employs 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.

[0062] In some examples, user device 104 communicates with DA server 106 via second user device 122. Second user device 122 is similar or identical to user device 104. For example, second user device 122 is similar to devices 200, 400, or 600 described below with reference to FIGs. 2A, 4, and 6A-B. User device 104 is 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 is configured to act as a proxy between user device 104 and DA server 106. For example, DA client 102 of user device 104 is 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 processes the information and return relevant data (e.g., data content responsive to the user request) to user device 104 via second user device 122.

[0063] In some examples, user device 104 is 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 is 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, in some examples, includes any number and type of user devices configured in this proxy configuration to communicate with DA server system 106.

[0064] Although the digital assistant shown in FIG. 1 includes 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 are 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 is 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

[0065] 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.

[0066] 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).

[0067] 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.

[0068] 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.

[0069] Memory 202 includes one or more computer-readable storage mediums. The computer-readable storage mediums are, for example, tangible and non-transitory. Memory 202 includes high-speed random access memory and also includes 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 controls access to memory 202 by other components of device 200.

[0070] In some examples, a non-transitory computer-readable storage medium of memory 202 is used to store instructions (e.g., for performing aspects of processes 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 the processes described below) are stored on a non-transitory computer-readable storage medium (not shown)

of the server system 108 or are divided between the non-transitory computer-readable storage medium of memory 202 and the non-transitory computer-readable storage medium of server system 108.

[0071] Peripherals interface 218 is 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 are implemented on a single chip, such as chip 204. In some other embodiments, they are implemented on separate chips.

[0072] 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.

[0073] 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 are 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).

[0074] 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).

[0075] A quick press of the push button disengages 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) turns power to device 200 on or off. The user is 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.

[0076] 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 includes graphics, text, icons, video, and any combination thereof (collectively termed "graphics"). In some embodiments, some or all of the visual output correspond to user-interface objects.

[0077] 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.

[0078] Touch screen 212 uses 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 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.

[0079] A touch-sensitive display in some embodiments of touch screen 212 is 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.

[0080] A touch-sensitive display in some embodiments of touch screen 212 is 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.

[0081] Touch screen 212 has, for example, a video resolution in excess of 100 dpi. In some embodiments, the touch screen has a video resolution of approximately 160 dpi. The user makes 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.

[0082] In some embodiments, in addition to the touch screen, device 200 includes 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 is a touch-sensitive surface that is separate from touch screen 212 or an extension of the touch-sensitive surface formed by the touch screen.

[0083] Device 200 also includes power system 262 for powering the various components. Power system 262 includes 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.

[0084] Device 200 also includes 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 includes 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 captures 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 is 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 is 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 is used along with the touch screen display for both video conferencing and still and/or video image acquisition.

[0085] 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.

[0086] Device 200 also includes one or more proximity sensors 266. FIG. 2A shows proximity sensor 266 coupled to peripherals interface 218. Alternately, proximity sensor 266 is coupled to input controller 260 in I/O subsystem 206. Proximity sensor 266 is performed 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).

[0087] 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.

[0088] Device 200 also includes one or more accelerometers 268. FIG. 2A shows accelerometer 268 coupled to peripherals interface 218. Alternately, accelerometer 268 is coupled to an input controller 260 in I/O subsystem 206. Accelerometer 268 performs, for example, 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.

[0089] 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 stores 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.

[0090] 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.

[0091] 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.

[0092] 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.

[0093] 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).

[0094] 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.

[0095] 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.

[0096] 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.

[0097] 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.

[0098] Text input module 234, which is, in some examples, a component of graphics module 232, provides soft keyboards for entering text in various applications (e.g., contacts 237, email 240, IM 241, browser 247, and any other application that needs text input).

[0099] 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).

[0100] Digital assistant client module 229 includes various client-side digital assistant instructions to provide the client-side functionalities of the digital assistant. For example, digital assistant client module 229 is 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 is also 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 is 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 communicates with DA server 106 using RF circuitry 208.

[0101] User data and models 231 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 include 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.

[0102] In some examples, digital assistant client module 229 utilizes 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 provides 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 also uses the contextual information to determine how to prepare and deliver outputs to the user. Contextual information is referred to as context data.

[0103] In some examples, the contextual information that accompanies the user input includes 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 is provided to DA server 106 as contextual information associated with a user input.

[0104] In some examples, the digital assistant client module 229 selectively provides 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 also elicits 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 passes 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.

[0105] 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.

[0106] Applications 236 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 includes, in some examples, 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.

[0107] Examples of other applications 236 that are 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.

[0108] In conjunction with touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, contacts module 237 are 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.

[0109] 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 are 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 uses any of a plurality of communications standards, protocols, and technologies.

[0110] 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.

[0111] 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.

[0112] 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 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).

[0113] 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.

[0114] 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.

[0115] 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.

[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, 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.

[0117] 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.

[0118] 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 can 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).

[0119] 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 are used by a user to create widgets (e.g., turning a user-specified portion of a web page into a widget).

[0120] 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.

[0121] 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.).

[0122] 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.

[0123] 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 are 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.

[0124] 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.

[0125] 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 can be combined or otherwise rearranged in various embodiments. For example, video player module can 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 stores a subset of the modules and data structures identified above. Furthermore, memory 202 stores additional modules and data structures not described above.

[0126] 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 is reduced.

[0127] 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.

[0128] 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).

[0129] 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.

[0130] 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.

[0131] 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.

[0132] 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).

[0133] In some embodiments, event sorter 270 also includes a hit view determination module 272 and/or an active event recognizer determination module 273.

[0134] 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.

[0135] 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 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 is called the hit view, and the set of events that are recognized as proper inputs is determined based, at least in part, on the hit view of the initial touch that begins a touch-based gesture.

[0136] 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.

[0137] 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.

[0138] 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.

[0139] 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.

[0140] 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 utilizes or calls 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.

[0141] 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 include sub-event delivery instructions).

[0142] 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 also includes 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.

[0143] 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.

[0144] 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.

[0145] 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.

[0146] 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.

[0147] 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 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.

[0148] 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.

[0149] 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.

[0150] 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.

[0151] 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.

[0152] 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.

[0153] 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.

[0154] Device 200 also includes one or more physical buttons, such as “home” or menu button 304. As described previously, menu button 304 is used to navigate to any application 236 in a set of applications that is 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.

[0155] 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.

[0156] 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.

[0157] Each of the above-identified elements in FIG. 4 is, in some examples, 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 are combined or otherwise rearranged in various embodiments. In some embodiments, memory 470 stores a subset of the modules and data structures identified above. Furthermore, memory 470 stores additional modules and data structures not described above.

[0158] Attention is now directed towards embodiments of user interfaces that can be implemented on, for example, portable multifunction device 200.

[0159] 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 are implemented on device 400. In some embodiments, user interface 500 includes the following elements, or a subset or superset thereof:

[0160] 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.

[0161] 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 is optionally 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.

[0162] 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.

[0163] 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.

[0164] 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.

[0165] FIG. 6A illustrates exemplary personal electronic device 600. Device 600 includes body 602. In some embodiments, device 600 includes 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) has 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) provide output data that represents the intensity of touches. The user interface of device 600 responds to touches based on their intensity, meaning that touches of different intensities can invoke different user interface operations on device 600.

[0166] Techniques for detecting and processing touch intensity are 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.

[0167] In some embodiments, device 600 has one or more input mechanisms 606 and 608. Input mechanisms 606 and 608, if included, are 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 permit device 600 to be worn by a user.

[0168] FIG. 6B depicts exemplary personal electronic device 600. In some embodiments, device 600 includes 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 is connected to display 604, which can have touch-sensitive component 622 and, optionally, touch-intensity sensitive component 624. In addition, I/O section 614 is 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 includes input mechanisms 606 and/or 608. Input mechanism 606 is a rotatable input device or a depressible and rotatable input device, for example. Input mechanism 608 is a button, in some examples.

[0169] Input mechanism 608 is a microphone, in some examples. Personal electronic device 600 includes, for example, 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 are operatively connected to I/O section 614.

[0170] Memory 618 of personal electronic device 600 is a non-transitory computer-readable storage medium, for storing computer-executable instructions, which, when executed by one or more computer processors 616, for example, cause the computer processors to perform the techniques and processes described below. The computer-executable instructions, for example, are also 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. 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.

[0171] As used here, the term "affordance" refers to a user-interactive graphical user interface object that is, for example, displayed on the display screen of devices 200, 400, 600, 810A-C, 820, 830, 840, 1182, 1186, 1880, and/or 1882 (FIGS. 2, 4, 6, 8A-8B, 9A-9C, 10A-10C, 11A-11D, 12A-12C, 13A-13B, 14, 15A-15G, and 18A-18E). For example, an image (e.g., icon), a button, and text (e.g., hyperlink) each constitutes an affordance.

[0172] 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).

[0173] 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 includes 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.

[0174] In some embodiments, a portion of a gesture is identified for purposes of determining a characteristic intensity. For example, a touch-sensitive surface receives 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 is 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 is 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.

[0175] The intensity of a contact on the touch-sensitive surface is 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.

[0176] 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.

[0177] 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).

[0178] 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).

[0179] 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

[0180] FIG. 7A illustrates a block diagram of digital assistant system 700 in accordance with various examples. In some examples, digital assistant system 700 is implemented on a standalone computer system. In some examples, digital assistant system 700 is distributed across

multiple computers. In some examples, some of the modules and functions of the digital assistant are 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, 600, 810A-C, 830, 840, 1182, 1186, 1880, and/or 1882) 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 is 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, can combine two or more components, or can have a different configuration or arrangement of the components. The various components shown in FIG. 7A are 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.

[0181] Digital assistant system 700 includes 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.

[0182] In some examples, memory 702 includes 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).

[0183] In some examples, I/O interface 706 couples 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, receives 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 includes any of the components and I/O communication interfaces described with respect to devices 200, 400, 600, 810A-C, 820, 830, 840, 1182, 1186, 1880, and/or 1882 in FIGS. 2, 4, 6A-6B, 8A-8B, 9A-9C, 10A-10C, 11A-11D, 12A-12C, 13A-13B, 14, 15A-15G, and 18A-18E, respectively. In some examples, digital assistant system 700 represents 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, 600, 810A-C, 830, 840, 1182, 1186, 1880, and/or 1882).

[0184] In some examples, the network communications interface 708 includes wired communication port(s) 712 and/or wireless transmission and reception circuitry 714. The wired communication port(s) receives and send communication signals via one or more wired interfaces, e.g., Ethernet, Universal Serial Bus (USB), FIREWIRE, etc. The wireless circuitry 714 receives and sends RF signals and/or optical signals from/to communications networks and other communications devices. The wireless communications 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 enables 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.

[0185] In some examples, memory 702, or the computer-readable storage media of memory 702, stores 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, stores instructions for performing the processes described below. One or more processors 704 execute these programs, modules, and instructions, and reads/writes from/to the data structures.

[0186] Operating system 718 (e.g., Darwin, RTXC, LINUX, UNIX, iOS, OS X, 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 communications between various hardware, firmware, and software components.

[0187] Communications module 720 facilitates communications between digital assistant system 700 with other devices over network communications interface 708. For example, communications module 720 communicates 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 also includes various components for handling data received by wireless circuitry 714 and/or wired communications port 712.

[0188] User interface module 722 receives 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 also prepares and delivers 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.).

[0189] Applications 724 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 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 include resource management applications, diagnostic applications, or scheduling applications, for example.

[0190] Memory 702 also stores digital assistant module 726 (or the server portion of a digital assistant). In some examples, digital assistant module 726 includes 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 has 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.

[0191] 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.

[0192] In some examples, as shown in FIG. 7B, I/O processing module 728 interacts 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 optionally obtains 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 includes 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 also sends follow-up questions to, and receives answers from, the user regarding the user request. When a user request is received by I/O processing module 728 and the user request includes speech input, I/O processing module 728 forwards the speech input to STT processing module 730 (or speech recognizer) for speech-to-text conversions.

[0193] STT processing module 730 includes 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 includes a front-end speech pre-processor. The front-end speech pre-processor extracts representative features from the speech input. For example, the front-end speech pre-processor performs 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 includes one or more speech recognition models (e.g., acoustic models and/or language models) and implements one or more speech recognition engines. Examples of speech recognition models include Hidden Markov Models, Gaussian-Mixture Models, Deep Neural Network Models, n-gram language models, and other statistical models. Examples of speech recognition engines 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 are 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 is 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 is passed to natural language processing module 732 for intent deduction. In some examples, STT processing module 730 produces multiple candidate text representations of the speech input. Each candidate text representation is a sequence of words or tokens corresponding to the speech input. In some examples, each candidate text representation is associated with a speech recognition confidence score. Based on the speech recognition confidence scores, STT processing module 730 ranks the candidate text representations and provides the n-best (e.g., n highest ranked) candidate text representation(s) to natural language processing module 732 for intent deduction, where n is a predetermined integer greater than zero. For example, in one example, only the highest ranked (n=1) candidate text representation is passed to natural language processing module 732 for intent deduction. In another example, the five highest ranked (n=5) candidate text representations are passed to natural language processing module 732 for intent deduction.

[0194] 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.

[0195] In some examples, STT processing module 730 includes and/or accesses a vocabulary of recognizable words via phonetic alphabet conversion module 731. Each vocabulary word is associated with one or more candidate pronunciations of the word represented in a speech recognition phonetic alphabet. In particular, the vocabulary of recognizable words includes a word that is associated with a plurality of candidate pronunciations. For example, the vocabulary includes the word "tomato" that is associated with the candidate pronunciations of /tə'meɪroʊ/ and /tə'mɑtoʊ/. Further, vocabulary words are associated with custom candidate pronunciations that are based on previous speech inputs from the user. Such custom candidate pronunciations are stored in STT processing module 730 and are associated with a particular user via the user's

profile on the device. In some examples, the candidate pronunciations for words are determined based on the spelling of the word and one or more linguistic and/or phonetic rules. In some examples, the candidate pronunciations are manually generated, e.g., based on known canonical pronunciations.

[0196] In some examples, the candidate pronunciations are ranked based on the commonness of the candidate pronunciation. For example, the candidate pronunciation /tə'meɪroʊ/ is ranked higher than /tə'mɑtoʊ/, 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 are ranked based on whether the candidate pronunciation is a custom candidate pronunciation associated with the user. For example, custom candidate pronunciations are 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 are associated with one or more speech characteristics, such as geographic origin, nationality, or ethnicity. For example, the candidate pronunciation /tə'meɪroʊ/ is associated with the United States, whereas the candidate pronunciation /tə'mɑtoʊ/ is associated with Great Britain. Further, the rank of the candidate pronunciation is 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'meɪroʊ/ (associated with the United States) is ranked higher than the candidate pronunciation /tə'mɑtoʊ/ (associated with Great Britain). In some examples, one of the ranked candidate pronunciations is selected as a predicted pronunciation (e.g., the most likely pronunciation).

[0197] When a speech input is received, STT processing module 730 is 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 first identifies the sequence of phonemes /tə'meɪroʊ/ 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."

[0198] In some examples, STT processing module 730 uses approximate matching techniques to determine words in an utterance. Thus, for example, the STT processing module 730 determines that the sequence of phonemes /tə'meɪroʊ/ corresponds to the word "tomato," even if that particular sequence of phonemes is not one of the candidate sequence of phonemes for that word.

[0199] Natural language processing module 732 ("natural language processor") of the digital assistant takes the n-best candidate text representation(s) ("word sequence(s)" or "token sequence(s)") generated by STT processing module 730, and attempts to associate each of the candidate text representations with one or more "actionable intents" recognized by the digital assistant. An "actionable intent" (or "user intent") represents 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 is 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 is 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, also dependents on the assistant's ability to infer the correct "actionable intent(s)" from the user request expressed in natural language.

[0200] In some examples, in addition to the sequence of words or tokens obtained from STT processing module 730, natural language processing module 732 also receives contextual information associated with the user request, e.g., from I/O processing module 728. The natural language processing module 732 optionally uses the contextual information to clarify, supplement, and/or further define the information contained in the candidate text representations received from STT processing module 730. The contextual information includes, 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 is, in some examples, dynamic, and changes with time, location, content of the dialogue, and other factors.

[0201] In some examples, the natural language processing is based on, e.g., ontology 760. Ontology 760 is 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" represents a task that the digital assistant is capable of performing, i.e., it is "actionable" or can be acted on. A "property" represents 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 defines how a parameter represented by the property node pertains to the task represented by the actionable intent node.

[0202] In some examples, ontology 760 is made up of actionable intent nodes and property nodes. Within ontology 760, each actionable intent node is linked to one or more property nodes either directly or through one or more intermediate property nodes. Similarly, each property node is 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 includes a "restaurant reservation" node (i.e., an actionable intent node). Property nodes "restaurant," "date/time" (for the reservation), and "party size" are each directly linked to the actionable intent node (i.e., the "restaurant reservation" node).

[0203] In addition, property nodes "cuisine," "price range," "phone number," and "location" are sub-nodes of the property node "restaurant," and are each 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 also includes a "set reminder" node (i.e., another actionable intent node). Property nodes "date/time" (for setting the reminder) and "subject" (for the reminder) are each linked to the "set reminder" node. Since the property "date/time" is relevant to both the task of making a restaurant reservation and the task of setting a reminder, the property node "date/time" is linked to both the "restaurant reservation" node and the "set reminder" node in ontology 760.

[0204] An actionable intent node, along with its linked concept nodes, is described as a "domain." In the present discussion, each domain is 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 includes 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 includes the actionable intent node "set reminder," and property nodes "subject" and "date/time." In some examples, ontology 760 is made up of many domains. Each domain shares one or more property nodes with one or more other domains. For example, the "date/time" property node is 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.

[0205] While FIG. 7C illustrates two example domains within ontology 760, other domains 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 is associated with a "send a message" actionable intent node, and further includes property nodes such as "recipient(s)," "message type," and "message body." The property node "recipient" is further defined, for example, by the sub-property nodes such as "recipient name" and "message address."

[0206] In some examples, ontology 760 includes all the domains (and hence actionable intents) that the digital assistant is capable of understanding and acting upon. In some examples, ontology 760 is modified, such as by adding or removing entire domains or nodes, or by modifying relationships between the nodes within the ontology 760.

[0207] In some examples, nodes associated with multiple related actionable intents are clustered under a "super domain" in ontology 760. For example, a "travel" super-domain includes a cluster of property nodes and actionable intent nodes related to travel. The actionable intent nodes related to travel includes "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) 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" share one or more of the property nodes "start location," "destination," "departure date/time," "arrival date/time," and "party size."

[0208] In some examples, each node in ontology 760 is 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 are the so-called "vocabulary" associated with the node. The respective set of words and/or phrases associated with each node are 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" includes 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" includes words and phrases such as "call," "phone," "dial," "ring," "call this number," "make a call to," and so on. The vocabulary index 744 optionally includes words and phrases in different languages.

[0209] Natural language processing module 732 receives the candidate text representations (e.g., text string(s) or token sequence(s)) from STT processing module 730, and for each candidate representation, determines what nodes are implicated by the words in the candidate text representation. In some examples, if a word or phrase in the candidate text representation is found to be associated with one or more nodes in ontology 760 (via vocabulary index 744), the word or phrase "triggers" or "activates" those nodes. Based on the quantity and/or relative importance of the activated nodes, natural language processing module 732 selects 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 is selected. In some examples, the domain having the highest confidence value (e.g., based on the relative importance of its various triggered nodes) is selected. In some examples, the domain is 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.

[0210] User data 748 includes 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 uses 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 is 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.

[0211] It should be recognized that in some examples, natural language processing module 732 is implemented using one or more machine learning mechanisms (e.g., neural networks). In particular, the one or more machine learning mechanisms are configured to receive a candidate text representation and contextual information associated with the candidate text representation. Based on the candidate text representation and the associated contextual information, the one or more machine learning mechanism are configured to determine intent confidence scores over a set of candidate actionable intents. Natural language processing module 732 can select one or more candidate actionable intents from the set of candidate actionable intents based on the determined intent confidence scores. In some examples, an ontology (e.g., ontology 760) is also used to select the one or more candidate actionable intents from the set of candidate actionable intents.

[0212] 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.

[0213] 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 generates a structured query to represent the identified actionable intent. In some examples, the structured query includes 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 says "Make me a dinner reservation at a sushi place at 7." In this case, natural language processing module 732 is 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 includes 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 generates 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} is not specified in the structured query based on the information currently available. In some examples, natural language processing module 732 populates 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 populates a {location} parameter in the structured query with GPS coordinates from the user device.

[0214] In some examples, natural language processing module 732 identifies multiple candidate actionable intents for each candidate text representation received from STT processing module 730. Further, in some examples, a respective structured query (partial or complete) is generated for each identified candidate actionable intent. Natural language processing module 732 determines an intent confidence score for each candidate actionable intent and ranks the candidate actionable intents based on the intent confidence scores. In some examples, natural language processing module 732 passes the generated structured query (or queries), including any completed parameters, to task flow processing module 736 ("task flow processor"). In some examples, the structured query (or queries) for the m-best (e.g., m highest ranked) candidate actionable intents are provided to task flow processing module 736, where m is a predetermined integer greater than zero. In some examples, the structured query (or queries) for the m-best candidate actionable intents are provided to task flow processing module 736 with the corresponding candidate text representation(s).

[0215] Other details of inferring a user intent based on multiple candidate actionable intents determined from multiple candidate text representations of a speech input are described in U.S. Utility Application Serial No. 14/298,725 for "System and Method for Inferring User Intent

From Speech Inputs," filed June 6, 2014, the entire disclosure of which is incorporated herein by reference.

[0216] Task flow processing module 736 is configured to receive the structured query (or queries) 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 are provided in task flow models 754. In some examples, task flow models 754 include procedures for obtaining additional information from the user and task flows for performing actions associated with the actionable intent.

[0217] As described above, in order to complete a structured query, task flow processing module 736 needs 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 invokes dialogue flow processing module 734 to engage in a dialogue with the user. In some examples, dialogue flow processing module 734 determines how (and/or when) to ask the user for the additional information and receives and processes the user responses. The questions are provided to and answers are received from the users through I/O processing module 728. In some examples, dialogue flow processing module 734 presents 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 generates 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 then populates 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.

[0218] Once task flow processing module 736 has completed the structured query for an actionable intent, task flow processing module 736 proceeds to perform the ultimate task associated with the actionable intent. Accordingly, task flow processing module 736 executes

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" includes 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 performs 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.

[0219] In some examples, task flow processing module 736 employs 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 acts 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 are specified by a respective service model among service models 756. Service processing module 738 accesses 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.

[0220] For example, if a restaurant has enabled an online reservation service, the restaurant submits 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 establishes 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.

[0221] In some examples, natural language processing module 732, dialogue flow processing module 734, and task flow processing module 736 are 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 is a dialogue response to the speech input that at least partially fulfills the user's intent. Further, in some examples, the generated response is output as a speech output. In these examples, the generated response is 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 is data content relevant to satisfying a user request in the speech input.

[0222] In examples where task flow processing module 736 receives multiple structured queries from natural language processing module 732, task flow processing module 736 initially processes the first structured query of the received structured queries to attempt to complete the first structured query and/or execute one or more tasks or actions represented by the first structured query. In some examples, the first structured query corresponds to the highest ranked actionable intent. In other examples, the first structured query is selected from the received structured queries based on a combination of the corresponding speech recognition confidence scores and the corresponding intent confidence scores. In some examples, if task flow processing module 736 encounters an error during processing of the first structured query (e.g., due to an inability to determine a necessary parameter), the task flow processing module 736 can proceed to select and process a second structured query of the received structured queries that corresponds to a lower ranked actionable intent. The second structured query is selected, for example, based on the speech recognition confidence score of the corresponding candidate text representation, the intent confidence score of the corresponding candidate actionable intent, a missing necessary parameter in the first structured query, or any combination thereof.

[0223] Speech synthesis module 740 is 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 is in the form of a text string. Speech synthesis module 740 converts the text string to an audible speech output. Speech synthesis module 740 uses 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 is configured to synthesize individual words based on phonemic strings corresponding to the words. For example, a phonemic string is associated with a word in the generated dialogue response. The phonemic string is stored in metadata associated with the word. Speech synthesis model 740 is 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 is performed on a remote device (e.g., the server system 108), and the synthesized speech is 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 is 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.

4. Exemplary functions of a digital assistant providing digital assistant services based on user inputs.

[0226] FIGs. 2, 4, 6, 8A-8B, 9A-9C, 10A-10C, 11A-11D, 12A-12C, 13A-13B, and 14 illustrate functionalities of providing digital assistant services by a digital assistant operating on an electronic device. In some examples, the digital assistant (e.g., digital assistant system 700) is implemented by a user device according to various examples. In some examples, the user device, a server (e.g., server 108, device 820), or a combination thereof, may implement a digital assistant system (e.g., digital assistant system 700). The user device can be implemented using, for example, device 200, 400, 600, 810A-C, 820, 830, 840, 1182, and/or 1186. In some

examples, the user device is a device having audio outputting capabilities and network connectivity, a smartphone, a laptop computer, a desktop computer, or a tablet computer.

[0227] FIGs. 8A-8B illustrate functionalities of providing digital assistant services at one or more electronic devices 810A-C based on a user input, according to various examples. In some examples, electronic device 810A (and similarly other electronic devices 810B-C) can include one or more audio input and output devices (e.g., a microphone and one or more speakers) and one or more network communication interfaces. Device 810A and devices 810B-C are collectively referred to as electronic devices 810 or devices 810. Device 810A, and similarly devices 810B-C, can include multiple speakers to provide surround sound. In some example, an electronic device 810 can further include one or more indicators (e.g., lights) for providing device operational indications. For example, one or more indicators of device 810A may emit light to indicate that device 810A is powered on, connected to a network, outputting audios, etc. Devices 810A-C can be service-extension devices to extend digital assistant services from other devices.

[0228] As illustrated in FIG. 8A, in some examples, the digital assistant operating on device 810A can be configured to communicatively couple to other electronic devices (e.g., devices 810B-C, 820, 830, and/or 840) via a direct communication connection, such as Bluetooth, near-field communication (NFC), BTLE (Bluetooth Low Energy), or the like, or via a wired or wireless network, such as a local Wi-Fi network. For example, a digital assistant operating on device 810A can detect devices 810B-C via Bluetooth discovery, and communicatively couple to devices 810B-C via a Bluetooth connection. As another example, the digital assistant operating on device 810A can detect a Wi-Fi network and communicatively couple to devices830 and 840 via a Wi-Fi network. As another example, the digital assistant operating on device 810A can detect a near field communication when the device 830 (e.g., a client device such as the user's smartphone) is in close proximity with, or physically in touch with, device 810A. For instance, to pair up device 810A and device 830, user 804 may tap device 810A with device 830, thereby establishing near-field communication between the two devices. As another example, the digital assistant operating on device 810A can detect that device 830 (e.g., a client device such as the user's smartphone) is within a predetermined distance (e.g., within a range of Bluetooth communication) and establish a connection with device 830. For instance, as user 804 approaches or enters area 800 with device 830, the digital assistant operating on device 810A

may detect that device 830 is within communication range, and thus connect with device 830. As another example, the digital assistant operating on device 810A can establish a connection with device 830 based on one or more previous established connections between the two devices. For instance, the digital assistant operating on device 810A can store a log file indicating the devices that it connected in the past, and optionally connection parameters. Thus, based on the log file, the digital assistant operating on device 810A can determine, for example, that it has connected to device 830 before. Based on such determination, the digital assistant operating on device 810A can establish the connection with device 830.

[0229] In some examples, electronic device 820 can be a server; and devices 830 and 840 can be client devices disposed in the vicinity of electronic devices 810. For example, device 820 can be a remotely disposed cloud server; and devices 830 and 840 can be the user's smartphone and a TV set-top box, respectively. In some examples, the digital assistant operating on device 810A establishes one or more connections with at least one of devices 820, 830, and 840, before it can receive user inputs comprising user requests and/or provide representations of user requests to one or more of devices 820, 830, and 840. Receiving user inputs and providing representations of user requests to devices 820, 830, and/or 840 are described in more detail below. Establishing connections before receiving user inputs and providing representations of user requests to other devices can improve the operation efficiency and speed of providing responses to user requests. For example, by establishing connections beforehand, the digital assistant operating on device 810A may not waste time to establish a connection after a user input is received.

[0230] In some examples, after establishing a connection between device 810A and device 830 (e.g., a client device such as the user's smartphone), the digital assistant operating on device 810A and/or device 830 can notify device 820 (e.g., a server) of the established connection. As described in more detail below, the digital assistant operating on device 810A may provide representations of a user request to one or both of device 820 and device 830 for obtaining responses. Device 820 may be a remote device such as a server and device 830 may be a device disposed in the vicinity of device 810A. Thus, notifying device 820 (e.g., a remote server) of a connection between device 810A and device 830 can facilitate an efficient operation. For example, as described below, in some embodiments, the digital assistant operating on device 810A may provide the representations of the user request to both device 820 and device 830.

Device 820 and/or device 830 may determine that device 830 (e.g., a client device disposed in the vicinity of device 810A) is capable of providing the response. Thus, because device 820 is notified that device 830 and device 810A are connected, device 820 may not provide a response to device 810A. Instead, device 820 may coordinate with device 830 to provide the response. In some examples, because device 830 (e.g., the user's smartphone) is disposed in the vicinity of device 810A, the response may be provided in a faster and more efficient manner.

[0231] In some examples, the digital assistant operating on devices 810A can establish connections with one or more devices having the same type. For example, as shown in FIG. 8A, a plurality of devices 810A-C can be service-extension devices and can be disposed in area 800. The digital assistant operating on device 810A can thus establish a connection with each of device 810B and device 810C. As described in more detail below, establishing connections between devices 810A, 810B, and 810C enables responses to be provided to user 804 by any device 810A-C disposed in area 800. This provides flexibility and improves the user-interaction efficiency. For example, user 804 may provide a speech input (e.g., "Play music") to device 810A, and receive a response by device 810C (e.g., music playing at device 810C).

[0232] In some embodiments, devices 810A-C can be service-extension devices for extending digital assistant services from one device to another. For example, as shown in FIG. 8A, devices 810A-C can be disposed in the vicinity of an electronic device 830 (e.g., a smartphone device) and/or electronic device 840 (e.g., a TV set-top box) to extend digital assistant services provided by electronic devices 830 and/or 840. In some examples, disposing devices 810A-C in the vicinity of devices 830 and/or 840 may include disposing devices 810A-C within a predetermined boundary surrounding, or a predetermined distance from, devices 830 and/or 840. For example, devices 810A-C may be disposed in the same house or building as devices 830 or 840. As shown in FIG. 8A, user 804 may be physically within or nearby an area 800, which may include one or more rooms 871, 873, and 875. User 804 may by physically located in room 871, while electronic device 830 (e.g., the user's smartphone) may be disposed in another room 873. In some examples, user 804 may want to access digital assistant services provided by device 830, despite that device 830 is incapable of directly communicating with user 804 (e.g., device 830 may not be able to directly receive the user 804's speech input via its microphone). In some examples, devices 810A-C can serve as service-extension devices for extending digital assistant services provided by device 830, as described in more detail below.

[0233] In some embodiments, one or more devices 810A-C may or may not be associated with a single device or user. Devices 810A-C (e.g., service-extension devices) can be shared by multiple users and can extend digital assistant services for multiple devices. In some examples, one or more devices 810A-C can extend digital assistant services to a plurality of users. As illustrated in FIG. 8B, user 804 and user 806 can share one or more devices 810A-C. For example, user 804 may have an associated device 830 (e.g., user 804's smartphone or smart watch); and user 806 may have an associated device 832 (e.g., user 806's smartphone or tablet). In some examples, the digital assistant operating on device 810A can establish a connection between itself and device 830; and a connection between itself and device 832. As such, the digital assistant operating on device 810A can extend digital assistant services for one or both devices 830 and 832. The capability of extending digital assistant services for multiple devices enables devices 810A-C to be shared by, for example, multiple users (e.g., family members).

[0234] With reference back to FIG. 8A, in some embodiments, a digital assistant operating on electronic device 810A can receive, from user 804, a speech input representing a user request. For example, user 804 may provide one or more speech inputs such as "What is on my calendar tomorrow?", "When is my first meeting?", "How is the weather?" or "Play Star Wars from my movie application." In some examples, the user request can be a request for information specific to user 804. For example, the speech inputs such as "What is on my calendar tomorrow?" or "When is my first meeting tomorrow?" represent requests for information specific to user 804. In some examples, the user request can be a request for non-user specific information. For examples, the speech inputs such as "How is the weather tomorrow?" or "What is the today's stock price of AAPL?" represent requests for information that is not specific to any particular user.

[0235] In some embodiments, prior to receiving the user's speech input, the digital assistant operating on device 810A can receive an additional speech input that includes a predetermined content. In response to receiving the additional speech input, the digital assistant operating on device 810A can activate device 810A. For example, device 810A may be placed in a standby mode or a lower power mode. Placing device 810A in standby mode or a lower power mode can reduce power consumption and in some examples, enhances the protection of the user's privacy. For example, during a standby mode or lower power mode, only limited voice detection and/or speech processing functions are enabled for the digital assistant operating on device 810A.

Other functions of device 810A (e.g., camera, indication light, speaker, etc.) may be disabled. In some examples, during the standby mode or lower power mode, the digital assistant operating on device 810A can still detect a speech input and determine whether the speech input includes predetermined content such as "Wake up, speaker" or "Hey, Speaker." Based on the determination, the digital assistant operating on device 810A can activate the device 810A. In some examples, after device 810A is activated, device 810A exits the standby mode and switches to a normal operation mode. In the normal operation mode, the digital assistant operating on device 810A can perform additional functions.

[0236] As illustrated in FIG. 8A, in an area 800 (e.g., a house), a plurality of devices 810A-C can be disposed. In some examples, a speech input activating one of device 810A-C may or may not activate other devices being disposed in the vicinity. For example, as described above, user 804 may provide a speech input including predetermined content (e.g., "Wake up, speaker") to device 810A. In some examples, device 810B may be disposed in another portion of area 800 (e.g., another room) and thus may not receive the speech input. As a result, device 810B may not be activated. In some examples, device 810B may be disposed in the vicinity of device 810A (e.g., in the same room) and may also receive the speech input including the predetermined content. In some examples, the digital assistant operating on device 810A can coordinate with device 810B to determine which device should be activated. For example, the digital assistants operating on device 810A and device 810B can both detect and record the volume or sound pressure associated with the speech input. Based on the comparison of the sound pressure detected at device 810A and the sound pressure detected at device 810B, the user's position relative to the two devices can be determined. For example, it may be determined that the user is physically closer to device 810A than to device 810B. As a result, device 810A may be activated while device 810B may not be activated. It is appreciated that the determination of which device is to be activated can be based on the user's speech input (e.g., user 804 provides "wake up, speaker in living room") and/or any context information such as the user's preferences, the user's relative position, the capabilities and attributes of the devices (e.g., one device is better for performing certain tasks than another device), etc.

[0237] In some embodiments, after receiving the user's speech input, the digital assistant operating on electronic device 810A can output one or more speech inquiries regarding the user request. For example, a user intent may not be determined or clear based on the user's speech

input; or the digital assistant operating on device 810A may not have properly received the speech input. As such, the digital assistant operating on device 810A may output a speech inquiry such as "What's that?" or "I did not quite get that," thereby seeking to clarify the user request. In response to the one or more speech inquiries, user 804 can thus provide one or more additional speech inputs clarifying his or her requests (e.g., repeat or rephrase the previous speech input). And device 810A can receive the one or more additional speech inputs.

[0238] In some embodiments, after receiving the user's speech input, the digital assistant operating on device 810A can obtain an identity of user 804. As described, electronic device 810A can extend digital assistant services for multiple devices (e.g., devices 830 and 832) associated with one or more users (e.g., users 804 and 806 as shown in FIG. 8B). Thus, for providing extension of digital assistant services from a proper device associated with a particular user (e.g., user 804 or user 806), the digital assistant operating on electronic device 810A can obtain the identity of the user. FIGs. 9A-9C illustrate functionalities of obtaining an identity of user 804 at electronic device 810A, according to various examples. With references to FIGs. 8A and 9A-9C, in some examples, electronic device 810A can include an authentication module 912. Authentication module 912 can include one or more sensors such as voice biometric sensors, facial recognition systems, fingerprint readers, NFC sensors, etc. In some examples, as shown in FIG. 9A, authentication module 912 can obtain authentication data 906 associated with user 804. In some examples, authentication data 906 can include user's voice biometrics, and/or fingerprint, and/or user's facial recognition data. For example, user 804's voice biometrics may include the user's voice characteristics such as acoustic patterns or voiceprints. User 804's facial recognition data may include the user's facial characteristics that may uniquely identify the user, such as the relative position, size, and/or shape of the eyes, nose, cheekbones, jaw, etc.

[0239] In some examples, authentication data 906 can include sensing of another electronic device that identifies the user. For example, the digital assistant operating on device 810A may detect that user 804's wearable device (e.g., a smart watch) is disposed in the vicinity of device 810A, communicate with the wearable device via NFC (e.g., Bluetooth), and obtain authentication data 906 from the wearable device (e.g., having already been authenticated on the user's watch). As another example, the digital assistant operating on device 810A may detect that device 810A is physically in contact with the user's smartphone that identifies the user, communicate with the user's smartphone via NFC (e.g., Bluetooth), and obtain authentication

data 906 from the user's smartphone. In some examples, authentication data 906 can include other credentials of the user, such as the user's fingerprints, passwords, or the like. It is appreciated that the digital assistant operating on electronic device 810A can obtain any authentication data associated with user 804 in any manner.

[0240] With reference to FIGs. 8A and 9A-9C, the digital assistant operating on device 810A can obtain a determination of the identity of the user 804 based on authentication data 906. As illustrated in FIG. 9B, in some embodiments, the digital assistant operating on electronic device 810A can provide the obtained authentication data 906 to electronic device 830 for authentication. For example, the digital assistant operating on device 810A can provide the user's voice biometrics data, the user's facial recognition data, the user's fingerprint data, device sensing data, and/or other credentials to electronic device 830. As described, electronic device 830 can be a device that is associated with user 804 (e.g., user 804's smartphone) and can thus store user identity information. Based on the received authentication data 906, device 830 can determine whether authentication data 906 include credentials that match with the stored user identity information (e.g., a password or a fingerprint). If authentication data 906 include credentials that match with user identity information, device 830 can send a determination of user identity 910 of user 804 to device 810A.

[0241] With reference to FIG. 9C, as described, the digital assistant operating on device 810A can provide authentication data 906 to device 830. In some embodiments, device 830 (e.g., a smartphone) may be incapable of obtaining the identity of user 804 and may thus forward authentication data 906 to electronic device 820. For example, device 830 may not store voice biometric information that can identify a user, and may thus be incapable of making a determination of user 804's identity. Device 830 may thus forward authentication data 906 to device 820. In some examples, device 820 may be disposed remotely from devices 810A and 830. For example, device 820 can be a server that is communicatively coupled to devices 810A and 830 via network(s) 850. Device 820 may store user identity information and may thus determine whether authentication data 906 include credentials that match with the stored identity information. If device 820 determines that authentication data 906 include credentials that match with user identity information of user 804, device 820 can send the determination of user identity 926 of user 804 to device 810A. In some examples, device 820 can send the determination of user identity 926 of user 804 directly to device 810A. In some examples, device 820 can send

the determination of user identity 926 of user 804 to device 830, which then forward to device 810A.

[0242] In some examples, obtaining the identity of user 804 can be based on a speech input including predetermined content. As described above, based on a speech input including predetermined content (e.g., "Wake up, speaker" or "Hey, Speaker"), the digital assistant operating on device 810A can activate device 810A. The speech input including the predetermined content can also be used to determine user 804's voice biometrics, which may include the user's voice characteristics such as acoustic patterns or voiceprints. As a result, a speech input including a predetermined content (e.g., a speech input for activating device 810A) can be used for identifying user 804, in manners similar to those described above.

[0243] FIGs. 10A-10C illustrate functionalities of providing digital assistant services based on a user request for information, according to various examples. With reference to FIGs. 8A and 10A-10C, in accordance with the obtained user identity, the digital assistant operating on electronic device 810A can provide a representation of the user request 1008 to at least one of device 820 or device 830. As described above, in some examples, device 820 may be a server disposed remotely from devices 810A-C and 830. Device 830 may be a client device associated with user 804 (e.g., the user's smartphone) and may be disposed in the vicinity of devices 810A-C (e.g., in the same house or building).

[0244] In some embodiments, the digital assistant operating on device 810A can provide the representation of the user request 1008 to a device that is disposed in the vicinity of device 810A before providing the representation of the user request to a remote device. As illustrated in FIG. 10A, user 804 may provide a speech input 1006 such as "When is my first meeting tomorrow?" Thus, speech input 1006 includes a user request for information regarding, for example, user 804's first meeting time on the next day. In some embodiments, the digital assistant operating on device 810A can determine whether device 830 is communicatively coupled to device 810A. For example, the digital assistant operating on device 810A can detect whether device 830 is within the range of communication and whether a connection can be established via NFC such as Bluetooth or Wi-Fi connections. In accordance with a determination that device 830 is communicatively coupled to device 810A, the digital assistant operating on device 810A can provide representation of the user request 1008 to device 830. In some embodiments, device 830

is a device that is disposed in the vicinity of device 810A (e.g., the user's smartphone); and the digital assistant operating on device 810A may not further provide representation of the user request to a remote device such as device 820 shown in FIG. 8A. As a result, the user request is not transmitted remotely, and may stay within a device that is disposed in the vicinity of device 810A (e.g., the user's personal devices). By providing the representation of the user request 1008 only to device 830, which is disposed in the vicinity of device 810A, a response from device 830 may be obtained in a fast and efficient manner without having to consume time for communication with a remote device. As a result, the speed of responding to a user request at device 810A may be improved. Moreover, a user request (e.g., the user requested included in speech input 1006) may include a request for sensitive or confidential user-specific information (e.g., the user's calendar information). As a result, for privacy concerns, it may be desired not to send the representation of the user request 1008 to a remote device, such as a cloud server.

[0245] As illustrated in FIG. 10A, in some example, device 830 receives the representation of user request 1008 from device 810A and determines whether it is capable of providing a response to the user request. For example, as described, the user request may include a request for information of user 804's first meeting time on the next day. Device 830 may determine that it stores user 804's calendar information is stored in device 830 and thus determine it is capable of providing the response to the user request. Accordingly, device 830 can send the response 1010 to the user request to device 810A. The response 1010 may include, for example, user 804's first meeting time on the next day. The digital assistant operating on device 810A receives the response 1010 to the user request from device 830, and can provide a representation of the response 1010 to user 804. As shown in FIG. 10A, the digital assistant operating on device 810A can provide a speech output 1012 such as "Your first meeting is at 9 a.m. tomorrow morning."

[0246] As described, the digital assistant operating on device 810A can determine whether device 830 is communicatively coupled to device 810A. For example, the digital assistant operating on device 810A can detect whether device 830 is within the range of communication and whether a connection between the two devices can be established via Bluetooth or Wi-Fi connections. With reference to FIG. 10B, in some embodiments, the digital assistant operating on device 810A can determine that device 830 is not communicatively coupled to device 810A For example, the digital assistant operating on device 810A may not be capable of detecting

device 830 because device 830 is beyond the range of communication, or because a connection cannot be established between the two devices. In accordance with a determination that the device 830 is not communicatively coupled to device 810A, the digital assistant operating on device 810A can provide the representation of the user request 1008 to device 820. As described above, device 820 can be a remote device such as a server. In some examples, the digital assistant operating on device 810A can provide the representation of the user request 1008 to device 820 via network(s) 850.

[0247] In some embodiments, as shown in FIG. 10B, device 820 receives the representation of user request 1008 from device 810A and determines whether it is capable of providing the response to the user request. For example, as described, the user request may include a request for information of user 804's first meeting time on the next day. Device 820 may determine that it stores, or has access to, user 804's calendar information (e.g., stored in user 804's cloud account) and thus determine it is capable of providing the response to the user request. Accordingly, device 820 can send a response 1014 to the user request to device 810A. The response 1014 to the user request may include, for example, the user's first meeting time on the next day. Device 810A receives the response 1014 to the user request from device 820, and can provide a representation of response 1014 to user 804. As shown in FIG. 10B, the digital assistant operating on device 810A can provide a speech output 1012 such as "Your first meeting is at 9 a.m. tomorrow morning." In some examples, after providing response to user 804, device 810A can continue to monitor subsequent speech inputs.

[0248] With reference to FIG. 10C, in some embodiments, user 804 may provide a speech input 1020 such as "What is the stock price of AAPL today?" This type of speech input represents a user request for non-user-specific information. Non-user-specific information is not specific to a particular user and may be general information such as weather information, stock price information, sports game information, etc. In some embodiments, as shown in FIG. 10C, the digital assistant operating on device 810A can provide the representation of user request 1022 for non-user-specific information to device 820 and not to device 830. As described, device 820 may be a server that is disposed remotely from device 810A, and device 830 may be the user's smartphone that is disposed in the vicinity of device 810A. In some embodiments, non-user-specific information (e.g., weather, stock price, game scores, etc.) may not be available and/or updated at device 830 (e.g., the user's smartphone). Thus, device 810A can determine that it is

more appropriate and efficient to obtain the non-user-specific information from a remote device (e.g., a server), rather than a device that is disposed in the vicinity of device 810A (e.g., a user's personal device). As such, the digital assistant operating on device 810A can provide the representation of user request 1022 for non-user-specific information to device 820 (e.g., a server) via network(s) 850.

[0249] As shown in FIG. 10C, device 820 receives the representation of user request 1022 from device 810A and determines that it is capable of providing the response to the user request. For example, as described, the user request may include a request for stock price information of AAPL. Device 820 may determine that it is capable of obtaining the information from a relevant data source (e.g., a finance website) and thus capable of providing the response to the user request. Accordingly, device 820 can send a response 1024 to the user request to device 810A. Response 1024 to the user request may include, for example, the current stock price of AAPL. The digital assistant operating on device 810A receives response 1024 from device 820, and can provide a representation of response 1024 to user 804. As shown in FIG. 10C, the digital assistant operating on device 810A can provide a speech output 1026 such as "AAPL closed at $200 today."

[0250] FIGs. 11A-11D illustrate functionalities of providing digital assistant services based on a user request for performing a task, according to various examples. With reference to FIGs. 11A and 11B, in some embodiments, user 804 may provide a speech input 1106 representing a user request for performing a task. For example, speech input 1106 may include "Play the Mighty Wings from Top Gun." Speech input 1106 thus represents a request to perform a task of playing a particular piece of music. In some examples, the digital assistant operating on device 810A may be incapable of (e.g., due to lack of sufficient information) determining whether a response to a user request can be provided by device 830 (e.g., a device disposed in the vicinity of device 810A such as the user 804's personal smartphone) or device 820 (e.g., a remote server). In the above example, device 810A may not have sufficient information to determine whether the user 804's smartphone or a server stores the song "Mighty Wings." Accordingly, the digital assistant operating on device 810A can provide the representation of user request 1108 to both device 830 and device 820.

[0251] As illustrated in FIG. 11A, device 820 and device 830 both receive the representation of the user request 1108 (e.g., a user request to perform a task). One or both of device 820 and device 830 can determine whether the respective device is capable of providing the response to the user request 1108. For example, device 830 can determine whether it stores the song Might Wings and if so, determine that it is capable of providing the response. Device 820 can make a similar determination. In some examples, the determinations can be made separately and independently on device 820 and device 830. For example, both devices 820 and 830 may determine whether they store the song "Mighty Wings;" and communicate the result of determination to the other device. In some examples, one of device 820 or device 830 can make the determination first and then send an indication to the other device. For example, device 830 may determine whether the song "Mighty Wings" is stored in device 830, and send an indication of the determination to device 820. If device 830 determines that it stores the song "Mighty Wings," it can send a corresponding indication to device 820 so that device 820 does not make any further determination. If device 830 determines that it does not have the song "Mighty Wings," it may send a corresponding indication to device 820 so that device 820 can then determine whether it stores, or has access to, the requested song. Similarly, device 820 may make a determination first and then send an indication to device 830. In some embodiments, the digital assistant operating on device 810A can cause one or both devices 820 and 830 to determine whether the respective device is capable of providing the response the user request. For example, the representation of the user request 1108 that is sent to devices 820 and 830 may include an explicit or implicit request for one or both devices 820 and 830 to determine whether one or both devices are capable of providing the requested response.

[0252] As shown in FIG. 11A, in some examples, device 820 (e.g., a server) may determine that it is capable of providing the response to the user request; and device 830 (e.g., the user's smartphone) may determine that it is incapable of providing the response to the user request. For example, device 820 may determine that it stores, or have access to, the requested song "Mighty Wings" and device 830 may determine that it does not store the song. Accordingly, device 820 can provide a response 1112 to the user request to device 810A. For example, device 820 can stream the song "Mighty Wings" to device 810A. Device 810A receives response 1112 from device 820 and provides a representation of response 1112 to user 804. For example, the digital

assistant operating on device 810A receives the streaming of the song "Mighty Wings" and provides audio outputs 1114 of the song.

[0253] With reference to FIG. 11B, in some examples, device 830 (e.g., the user's smartphone) may determine that it is capable of providing the response to the user request; and device 820 (e.g., the server) may determine that it is incapable of providing the response to the user request. For example, device 830 may determine that it stores the song "Mighty Wings" and device 820 may determine that it does not store the song, or does not have access, to the requested song without requiring a further user interaction (e.g., asking the user to purchase the song). Accordingly, device 830 can provide a response 1112 to the user request to device 810A. For example, device 830 can stream the song "Mighty Wings" to device 810A. The digital assistant operating on device 810A receives response 1122 from device 830 and provides a representation of response 1122 to user 804. For example, device 810A receives the streaming of the song "Mighty Wings" and provides audio outputs 1124 of the song.

[0254] With reference to FIGs. 11A and 11B, in some examples, both device 830 (e.g., the user's smartphone) and device 820 (e.g., a server) may determine that the respective device is capable of providing the response to the user request. For example, device 830 may determine that it stores the song "Mighty Wings;" and device 820 may determine that it also stores the requested song (e.g., in the user's cloud account) or has access to the song without requiring further user interaction (e.g., without requiring the user to purchase the song). Accordingly, either device 820 or device 830 is capable of providing a response to the user request to device 810A. In some examples, the selection of a device from multiple devices for providing the response can be based on a predetermined condition. For example, the predetermine condition may include a pre-configured policy (e.g., device 830 is the default device to provide response if more than one device are capable of providing responses), a condition of connection bandwidth (e.g., the device that has a higher bandwidth of connection is the device to provide response), a condition of user's preferences (e.g., in order to save cellular data usage, the user prefers to use a device that is connected to device 810A via Wi-Fi for providing responses), or the like. Based on the predetermine condition, one of device 820 and device 830 can stream the song "Mighty Wings" to device 810A. The digital assistant operating on device 810A receives the response to the user request and provides a representation of the response to the user. For example, device

810A receives the streaming of the song "Mighty Wings" and provides audio outputs of the song.

[0255] With reference to FIG. 11C, in some embodiments, user 804 may provide a speech input 1126 representing a user request for performing a task. Speech input 1126 may include, for example, "Play the movie Star Wars." The digital assistant operating on device 810A receives speech input 1126. In some examples, based on speech input 1126, device 810A can provide a representation of a user request 1128 to a device that is disposed in the vicinity of device 810A (e.g., device 840) and not to a remote device (e.g., device 820 such as a server). The digital assistant operating on device 810A may not provide the representation of user request 1128 to a remote device for a number of reasons. For example, the digital assistant operating on device 810A may determine that the information is likely to be available at a device that is disposed in the vicinity of device 810A (e.g., device 840 such as a TV set-top box); that there is no or poor connection to a remote device; that the bandwidth to a remote device is limited or inferior; that a predetermined configuration requires providing representations of user requests to a device that is the vicinity of device 810A (e.g., a device connected to device 810A via Wi-Fi); or the like. As described above, device 840 may be a TV set-top box that is disposed in the vicinity of device 810A, and device 820 may be a server disposed remotely. In some examples, the digital assistant operating on device 810A can be configured to always provide the representation of a user request to a device that is disposed in the vicinity of device 810A (e.g., device 840). In some examples, the digital assistant operating on device 810A can be configured to provide the representation of a user request to a device that is disposed in the vicinity of device 810A (e.g., device 840) or a remote device (e.g., device 820) based on the type and/or content of the user request. As described above, in some examples, if the user request is a request for user-specific information, the digital assistant operating on device 810A can provide the representation of the user request to a device that is disposed in the vicinity of device 810A (e.g., the user's smartphone); and if the user request is a request for non-user-specific information, device 810A can provide the representation of the user request to a remote device (e.g., the server).

[0256] As shown in FIG. 11C, device 840 receives the representation of the user request 1128, which may cause device 840 to determine whether it is capable of providing a response to the user request. For example, device 840 may determine that it stores the movie Star Wars and thus is capable of providing the response to device 810A. As other examples, device 840 may

determine it stores data including user's personal calendar, contacts, photos, media items, or the like, and thus is capable of providing a response for a user request for information or task performance using these stored data. In accordance with a determination that device 840 is capable of providing the response to the user request, device 840 can provide a response 1134 to device 810A. For example, device 840 can stream the movie "Star Wars" to device 810A. The digital assistant operating on device 810A receives response 1134 from device 810A, and provides a representation of response 1134 to the user. For example, the digital assistant operating on device 810A can provide audio outputs 1136 (e.g., play the movie "Star Wars") using its display and speakers. In some examples, device 840 can provide at least a portion of the response to device 810A while provide other portions of the response to one or more other devices. For example, device 840 can provide the audio portion of the movie "Star Wars" to device 810A while provide the video portion of the movie to a device 1137 (e.g., a TV).

[0257] With reference to FIG. 11C, as described, device 840 receives the representation of the user request 1128, which may cause device 840 to determine whether it is capable of providing the response to the user request. In some examples, device 840 may determine that it is incapable of providing the response to the user request. For example, device 840 may determine that it does not store the movie "Star Wars" and thus cannot provide the response. As other examples, device 840 may determine that the user request is for information that is not stored in device 840 (e.g., stock information, web searching request, etc.), and thus determine it cannot provide the response.

[0258] As shown in FIG. 11C, in accordance with a determination that the device 840 is incapable of providing a response to the user request, device 840 can forward a representation of user request 1128 to device 820 via network(s) 850. As described, device 820 can be a server. Based on the representation of user request 1128, device 820 can then determine whether it is capable of providing a response. For example, device 820 may determine whether it stored a requested movie, or whether a requested movie is accessible from user 804's cloud account or from a web source (e.g., a media website). If device 820 determines that it stores the requested movie or that the requested movie is accessible, device 820 determines that it is capable of providing the response. In some examples, device 820 can provide a response 1132 to device 840, which can then forward to device 810A and optionally device 1137 (e.g., a TV). In some examples, device 820 can provide response 1132 directly to device 810A and optionally device

1137. For example, device 820 can send the audio portion of the movie Star Wars to device 810A, while sending the video portion of the movie Star Wars to device 1137 (via device 840). The digital assistant operating on device 810A, and optionally device 1137, receives response 1132, and provides a representation of response 1132 to the user. For example, the digital assistant operating on device 810A, and optionally device 1137, can provide output 1136 based on the received response 1132 (e.g., play the movie "Star Wars").

[0259] Using the example illustrated in FIG. 11C, instead of providing speech input 1126, user 804 may provide a speech input such as "Play the movie Star Wars on my TV and set up another screen on my computer." Similar to those described above, the digital assistant operating on device 810A can provide a representation of the user request to device 840, and receive a response to the user request from device 840 or device 820, based on a determination whether device 840 (e.g., a TV set-top box disposed in the vicinity of device 810A) or device 820 (e.g., a server) is capable of providing the response. In some embodiments, the user request may indicate that the response to the user request is to be provided to multiple devices. Thus, device 840 and/or device 820 can provide the response accordingly. For example, the digital assistant operating on device 810A can receive a portion of the response (e.g., audio portion of a movie); device 1137 can receive another portion of the response (e.g., the video portion of a movie); and another device (e.g., the user's computer) can receive a duplicate copy of the response (e.g., a copy of both the audio and video portions of the movie). In some examples, user 804 may want to watch a movie using a device such as his or her computer or tablet, but not using device 810A. User 804 may provide a speech input such as "Play the movie Star Wars on my computer" or "Play the movie Star Wars on my tablet." The speech input may be provided as an initial input to start the task performance (e.g., start to play the movie Star Wars). The speech input may also be provided as a subsequent input while a task is being performed (e.g., while device 840 is streaming the movie to device 810A and/or device 1137). Similar to those described above, device 810A can provide a representation of the user request to device 840 and/or device 820. The representation of the user request may indicate a response is to be provided to the user's computer or tablet (not shown in FIG. 11C). The user's computer or tablet can thus receive a response to the user request from device 840 or device 820, based on a determination whether device 840 (e.g., a TV set-top box disposed in the vicinity of device 810A) or device 820 (e.g., a server) is capable of providing the response.

[0260] With reference to FIG. 11D, in some embodiments, user 804 may provide a speech input 1152 such as "Call Jane and conference Kevin." The digital assistant operating on device 810A receives speech input 1152 and can provide a representation of a user request 1154 to device 830 (e.g., the user's smartphone). User request 1154 can include a request to perform a task at device 830 (e.g., calling Jane and conferencing Kevin). Device 830 receives the representation of user request 1154, and determines that it is capable of performing the task. As described above in connection with FIGs. 1-7C, and similarly in other examples, a natural language processing module of the digital assistant operating on device 830 (and/or device 810A) can identify an actionable intent based on the user request and generate a structure query to represent the identified actionable intent. For example, based on speech input 1152, device 830 can thus determine if the actionable intent is "making phone calls." In some examples, the digital assistant can actively elicit and obtain information needed to fully infer the user intent (e.g., by disambiguating words, elicit further clarification inputs from the user, and/or use context information such as the user's contact list). A structured query for "making phone calls" may include parameters such as {callees}, {telephone numbers}, and the like. Next, a task flow processing module of the digital assistant can receive the structured query and perform the actions required to provide a response to the user request. Accordingly, device 830 can perform the task according to user request 1154 (e.g., call user 1194's device 1182 and conference in user 1196's device 1186). Based on the performance of the task, device 830 can also provide a response 1157 to device 810A. For example, the digital assistant operating on device 810A can receive response 1157 from device 830, indicating that the conference with user 1194 (e.g., Jane) and user 1196 (e.g., Kevin) has been established. Accordingly, the digital assistant operating on device 810A can provide an audio output 1162 such as "Jane and Kevin are connected."

[0261] FIGs. 12A-12C illustrate functionalities of providing digital assistant services based on a user request for information, according to various examples. With reference to FIG. 12A, user 804 may provide a speech input 1206 such as "Find Jane's mobile phone number." Speech input 1206 thus represents a user request for a telephone number. The digital assistant operating on device 810A receives speech input 1206, and provides a representation of a user request 1208 to a remote device (e.g., device 820 such as a server) via network(s) 850, and not to a device that is disposed in the vicinity of device 810A (e.g. device 830 such as the user's smartphone). The digital assistant operating on device 810A may not provide the representation of user request

1208 to a device that is disposed in the vicinity of device 810A for a number of reasons. For example, the digital assistant operating on device 810A may determine that the information is unlikely to be available at a device that is disposed in the vicinity of device 810A (e.g., device 830); that there is no or poor connection to a device that is in the vicinity of device 810A (e.g., device 830 is out of communication range with device 810A); that the bandwidth to a device that is disposed in the vicinity of device 810A is limited or inferior; that a predetermined configuration requires providing representations of user requests to a remote device; or the like.

[0262] In some embodiments, device 820 receives the representation of user request 1208, which may cause device 820 to determine whether it is capable of providing the response to the user request. For example, device 820 may determine that the user 804's cloud account stores the requested telephone number and thus is capable of providing a response to the user request. In accordance with a determination that device 820 is capable of providing a response to the user request, device 820 can provide response 1210 to device 810A. For example, device 820 can provide Jane's telephone number to device 810A. The digital assistant operating on device 810A receives response 1210 from device 820, and provides a representation of response 1210 to user 804. For example, the digital assistant operating on device 810A can provide an audio output 1212 such as "Jane's number is 123-456-7890."

[0263] With reference to FIG. 12B, similar to FIG. 12A, after device 810A receives a speech input 1226 such as "Find Jane's phone number," it can provide a representation of a user request 1228 to device 820 via network(s) 850. In some embodiments, device 820 receives the representation of user request 1228, which may cause device 820 to determine whether it is capable of providing a response to user request 1228. For example, device 820 may determine that the user 804's cloud account does not store Jane's telephone number and thus is incapable of providing a response to the user request. In accordance with a determination that device 820 is incapable of providing the response to the user request, device 820 can forward the representation of user request 1228 to device 830. Device 830 can be a device disposed in the vicinity of device 810A and can be a device associated with user 804 (e.g., user 804's personal device such as a smartphone). Similar to those described above, device 830 can determine whether it is capable of providing a response to the user request (e.g., whether it stores Jane's telephone number), and provides response 1232 to device 810A in accordance with the determination. For example, in accordance with a determination that device 830 is capable of

providing Jane's telephone number, device 830 can provide Jane's telephone number to device 810A. The digital assistant operating on device 810A receives the response 1232 from device 830, and provides a representation of response 1232 to the user. For example, the digital assistant operating on device 810A can provide an audio output 1234 such as "Jane's number is 123-456-7890." In some examples, the digital assistant operating on device 810A can receive the response directly from device 830. In some embodiments, device 830 can provide a response to device 820, which then forward the response to device 810A, as described below.

[0264] With reference to FIG. 12C and continuing the above example described in connection with FIG. 12B, the digital assistant operating on device 810A can receive a response 1252 indirectly from device 830. For example, device 830 can provide a response 1252 (e.g. Jane's telephone number) to device 820 (e.g., a server), which can then forward response 1252 to device 810A. The digital assistant operating on device 810A receives response 1252 from device 820, and provides a representation of response 1252 to user 804. For example, the digital assistant operating on device 810A can provide an audio output 1256 such as "Jane's number is 123-456-7890."

[0265] FIGs. 13A-13B illustrate functionalities of providing digital assistant services at a first electronic device or additional electronic devices, according to various examples. With reference to FIG. 13A, as described above, a plurality of devices 810A-C can be service-extension devices for extending digital assistant services from one device to another. For example, as shown in FIG. 13A, devices 810A-C can be disposed in the vicinity of device 830 (e.g., user 804's smartphone device) to extend digital assistant services provided by device 830. In some examples, disposing a plurality of devices 810A-C in the vicinity of device 830 may include disposing devices 810A-C within a predetermined boundary or distance of device 830. For example, devices 810A-C may be disposed in the same house or building as device 830. As shown in FIG. 13A, in some embodiments, devices 810A-C may be disposed in a manner for extending digital assistant services to different portions of an area 1300. As shown in FIG. 13, area 1300 may include, for example, a living room 1320, an office 1340, and a bedroom 1360. In some examples, device 810A can be disposed in living room 1320, device 810B can be disposed in office 1340; and device 810C can be disposed in bedroom 1360. As described above, devices 810A-C can be communicatively coupled to each other and to other devices (e.g., devices 820 and 830).

[0266] As shown in FIG. 13A, user 804 may be located within living room 1302, in which device 810A is disposed. User 804 may want to go to bed with some light music, and thus provide a speech input 1306 such as "Play light music on my bedroom speaker" to device 810A. The digital assistant operating on device 810A receives speech input 1306 representing a user request. Similar to those described above, device 810A can provide a representation of the user request to at least one of device 830 (e.g., the user's smartphone disposed in the vicinity of device 810A) or device 820 (e.g., a remoted disposed server). At least one of device 820 or device 830 determines whether the respective device is capable of providing a response to the user request, and provides the response to device 810A. The digital assistant operating on device 810A can then provide a representation of the response to user 804 (e.g., an audio output).

[0267] In some embodiments, prior to provide a representation of the response to user 804, the digital assistant operating on device 810A can determine whether the representation of the response is to be provided by device 810A or another device. For example, speech input 1306 may include "Play light music on my bedroom speaker." Accordingly, the digital assistant operating on device 810A can determine that the user intent is not to play the music on device 810A, but rather on device 810C disposed within bedroom 1360. The determination can be made using, for example, natural language processing described above. In accordance with a determination that the representation of the response is not to be provided by device 810A, device 810A can forward the response to, or cause the response to be provided to, device 810C disposed in bedroom 1360. Device 810C can thus provide an audio output 1310 playing the light music user requested. In other examples, in accordance with a determination that the representation of the response is to be provided by device 810A, device 810A can itself provide the representation of response to the user 804. As described above, multiple devices 810A-C can be disposed in area 1400. In some examples, a digital assistant (the digital assistant operating on device 810A, 810B, 810C, device 830, etc.) can determine the location of each device 810A-C (e.g., based on an initial configuration). The details of the location determination, as an illustrative example, are described in co-pending U.S. Patent Application entitled "WHOLE HOME AUDIO CONTROL INTERFACE," filed on May 16, 2017 (Attorney Docket No. 77000-30167.00 (P34482USP1)), the content of which is hereby incorporated by reference in its entirety, and included in the .

[0268] As described above, prior to provide a representation of the response to user 804, the digital assistant operating on device 810A can determine whether a representation of the response is to be provided by device 810A or another device. In some examples, such determination can be based on the user request represented by the user's speech input (e.g., "Play light music on my bedroom speaker"). In some examples, such determination can be based on at least one of detecting a location of the user or tracking the user's movement. With reference to FIG. 13B, user 804 may be located within living room 1302, in which device 810A is disposed. User 804 may want to go to office 1340 and having some light music, and thus provide a speech input 1326 such as "Play light music" to device 810A. Speech input 1326 does not indicate on which device 810A-C user would like the music to be played. The digital assistant operating on device 810A receives speech input 1326 representing a user request. Similar to those described above, device 810A can provide a representation of the user request to at least one of device 830 (e.g., the user's smartphone disposed in the vicinity of device 810A) or device 820 (e.g., a remoted disposed server). At least one of device 820 or device 830 determines whether the respective device is capable of providing a response to the user request, and provides the response to device 810A.

[0269] Prior to provide a representation of the response to user 804, device 810A can determine whether the representation of the response (e.g., an audio output) is to be provided by device 810A or another device. In some examples, device 810A can make such determination based on at least one of detecting of user 804's location or tracking user 804's movement. For example, device 810A may detect that user 804 is located in living room 1320 but is moving toward office 1340. Device 810A can detect location and/or movement using, for example, one or more sensors such as motion sensors, positioning systems, cameras, etc. In accordance with a determination that user 804 is moving toward office 1340, the digital assistant operating on device 810A can determine that the user intent is not to play the music on device 810A, but rather on device 810B disposed within office 1340, or that the music playback should be started on device 810A, but continued on device 810B disposed in office 1340 (and optionally discontinued on device 810A). In accordance with a determination that the representation of the response is not to be provided by device 810A, device 810A can forward the response, or cause the response to be provided, to device 810B disposed in office 1340. Device 810B can thus provide an audio output 1328 playing the light music user requested.

[0270] In other examples, in accordance with a determination that the representation of the response is to be provided by device 810A (e.g., user 804 is located in living room 1320 and not moving), device 810A can itself provide the representation of response to the user 804. It is appreciated that the digital assistant operating on device 810A can determine whether the response is to be provided device 810A or another device based on any context information, such as the user's preferences (e.g., user 804 prefers to listen to music before bedtime), past devices used for providing responses, device attributes and capabilities (e.g., device 810A may provide better sound than device 810B), etc.

[0271] FIG. 14 illustrates functionalities of providing continuity of digital assistant services between different electronic devices, according to various examples. As illustrated in FIG. 14, the digital assistant operating on device 810A may be providing a response 1406 to user 804 (e.g., playing music). While device 810A is in the process of providing response 1406, user 804 may move out of area 1400, in which device 810A is disposed. For example, user 804 may need to leave his or her house and go to work. In some embodiments, the digital assistant operating on device 810A can determine whether the response 1406 is to be continually provided a different electronic device. As an example, while device 810A is providing response 1406, user 804 may provide a speech input such as "Continue to play the music on my smartphone." Device 810A receives this speech input and can determine that the user intent is to continue to play the music on device 830 (e.g., the user's smartphone). Such determination can be made using natural language processing techniques described above. Based on the determined user intent, device 810A can determine that response 1406 should be to continue to provide the music at a different device.

[0272] In some embodiments, user 804 can also provide the speech input such as "Continue to play the music on my smartphone" to device 830, instead to device 810A. Based on the speech input and context information (e.g., device 810A is currently providing audio outputs), device 830 can determine the user intent is to continually perform a task that is being performed at device 810A. For example, device 830 can communicate with device 810A (and other devices communicatively coupled to device 830) to determine the status information of device 810A. The status information of device 810A may indicate that it is currently playing music. Accordingly, device 830 may determine that the user intent is to continually play the music that is being currently played on device 810A. Based on the determination, device 830 can

communicate with device 810A for continually performing the task that is being currently performing by device 810A. For example, device 830 can obtain the content and/or metadata (e.g., time stamps associated with the currently playing music), continually play the music by device 830, and cause device 810A to stop playing.

[0273] As another example illustrated in FIG. 14, while the digital assistant operating on device 810A is providing response 1406, device 810A can perform at least one of detecting a location of the user or tracking the user's movement. Device 810A can detect location and/or movement using, for example, one or more sensors such as motion sensors, positioning systems, cameras, etc. As an example, device 810A can continuously or periodically track the user 804's current location and/or movement. In some examples, device 810A can detect whether the user 804's location variation with respect to device 810A satisfies a predetermined condition. For example, device 810A can detect the user 804 has moved out of a predetermined boundary of area 1400 (e.g., a house). As a result, device 810A can determine that response 1406 should be continually provided at a different device.

[0274] As another example, while device 810A is in the process of providing response 1406, it can detect the movement of a device associated with user 804 (e.g., device 830 such as the user's smartphone). For example, device 810A can determine that the communication signal strength of device 830 reduced over a short duration of time, indicating that device 830 likely moves out of the boundary of area 1400. As a result, device 810A can determine that response 1406 should be continually provided at a different device (e.g., device 830).

[0275] In some embodiments, in accordance with a determination that response 1406 is to be continually provided at a different electronic device, device 810A can cause response 1406 to be continually provided by one or more different electronic devices. For example, device 810A can transmit the remaining content (e.g., the rest of the response 1406) for providing response 1406 and/or metadata associated with providing the response 1406 (e.g., the timestamp of the current playing media that was streamed from device 820 or device 830) to device 830 (e.g., the user's smartphone). In some examples, device 810A can also send a notification to another device (e.g., device 820), from which the content of response 1406 is obtained. The notification may indicate or request that the response 1406 is to be continually provided at another device and thus the content of response 1406 should be provided to that device. Based on the received remaining

content and/or metadata, device 830 can continue to provide response 1406 to user 804. More details of continually providing digital assistant services on a different device are described in co-pending U.S. Patent Application No. 15/271,766, entitled "INTELLIGENT DIGITAL ASSISTANT IN A MULTI-TASKING ENVIRONEMENT," filed September 21, 2016, the content of which is hereby incorporated by reference in its entirety, and included in the .

[0276] In the above description in connection with FIGs. 8A-8B, 9A-9C, 10A-10C, 11A-11D, 12A-12C, 13A-13B, and 14, device 820 can be a remote device such as a server. In some embodiments, a device can be disposed in the vicinity of devices 810A-C, operating as a proxy device for device 820. As one example and with reference back to FIG. 8A, device 840 (e.g., a TV set-top box) can operate as a proxy device for device 820 (e.g., a remote server). A proxy device can operate as an intermediary for requests from client devices (e.g., device 810A) seeking resources from other devices (e.g., servers). As a proxy, device 840 can operate to process requests from a plurality of home automation devices (e.g., a smart thermostat, a smart door, a smart light switch, etc.). For example, based on user's speech inputs (e.g., speech input received via device 810A), a smart thermostat may be required to perform a task of adjusting temperature and/or humidity levels. The smart thermostat may thus communicate with device 840 to request current temperature and humidity data from various sensors. Device 840 can thus operate as a proxy to relay the request to appropriate devices and/or sensors and provide the data to the smart thermostat.

5. Exemplary functions of a digital assistant providing digital assistant services based on notifications of events.

[0277] FIGs. 2, 4, 6, and 15A-15G illustrate functionalities of providing digital assistant services by a digital assistant operating on an electronic device. In some examples, the digital assistant (e.g., digital assistant system 700) is implemented by a user device according to various examples. In some examples, the user device, a server (e.g., server 108, device 820), or a combination thereof, may implement a digital assistant system (e.g., digital assistant system 700). The user device can be implemented using, for example, device 200, 400, 600, 810A-C, 820, and/or 830. In some examples, the user device is a device having audio outputting capabilities and network connectivity, a smartphone, a laptop computer, a desktop computer, or a tablet computer.

[0278] FIGs. 15A-15G illustrate functionalities of providing digital assistant services based on notifications of events, according to various examples. As illustrated in FIG. 15A, device 810A can receive a notification 1506 and/or 1508 of one or more events associated with user 1504. As described above, device 810A (and similarly other devices 810B-C) can include one or more audio input and output devices (e.g., a microphone and one or more speakers), one or more network communication interfaces, and optionally one or more indicators (e.g., lights) for providing device operational indications. In some examples, as shown in FIG. 15A, device 810A may receive notification 1506 from device 830 (e.g., the user's smartphone) and/or notification 1508 from device 820 (e.g., a remote server).

[0279] In some examples, a notification of an event can include a representation of at least one of an incoming call, a reminder, a message, a voicemail, a news alert, or the like. For example, a digital assistant operating on device 830 may receive a calendar reminder from a calendar application, and may forward a representation of the calendar reminder to device 810A. As shown in FIG. 15A, in response to receiving the notification 1506 and/or 1508, the digital assistant operating on device 810A can output one or more indications 1510 of notification 1506 and/or 1508. In some examples, an indication 1510 can be an audio indication (e.g., a beep, a tone, etc.), a visual indication (e.g., flashing lights, displayed messages, etc.), or a combination of audio and visual indications. While FIG. 15A illustrates that indication 1510 is provided by device 810A, indication 1510 can also be provided by other devices, such as device 830 and/or devices 810B-C (not shown in FIG. 15A). As described, device 830 can be, for example, the user's smartphone, smart watch, tablet, etc.; and devices 810B-C can be similar type of devices as device 810A and are disposed in the vicinity of device 810A (e.g., in the same house). Thus, an indication of a notification can be provided to user 1504 by any device at any location. Providing indications by multiple devices disposed at various locations can improve the likelihood of capturing user 1504's attention regarding the notifications. In some examples, a notification is only provided by one device (e.g., device 810A) in order to minimize the disturbance to user 1504.

[0280] As illustrated in FIG. 15B and continuing the above example, user 1504 receives indication 1510 and may provide a speech input 1516 inquiring about indication 1510. For example, speech input 1516 can include "What is it?" In some examples, the digital assistant operating on device 810A receives speech input 1516 and can output a response 1518 in

accordance with the notification of the event. For example, as shown in FIG. 15B, if the notification of the event includes a representation of a voice message from John, the digital assistant operating on device 810A can output a response 1518 such as "You have a voicemail from John." If the notification of the event includes a representation of a calendar reminder, the digital assistant operating on device 810A can output a response 1518 such as "You have an upcoming event on your calendar." If the notification of the event includes a representation of an incoming call from John, the digital assistant operating on device 810A can output a response 1518 such as "You have an incoming call from John."

[0281] As illustrated in FIG. 15C and continuing the above example, after outputting response 1518 in accordance with the notification of the event, the digital assistant operating on device 810A can continue to monitor user inputs by, for example, listening for user utterances during or after response 1518. For example, the digital assistant operating on device 810A may receive a subsequent speech input 1526. Speech input 1526 may include, for example, "Play the message," "What is the event?", or "Take the call from John."

[0282] In some examples, as shown in FIG. 15C, the digital assistant operating on device 810A receives speech input 1526 and can determine the user intent based on speech input 1526. For example, the digital assistant operating on device 810A can determine that the user intent is to play the voicemail from John, listen to the upcoming calendar event, or take the call from John. Accordingly, the digital assistant operating on device 810A can provide the notification to user 1504 in accordance with the determined user intent. For example, the digital assistant operating on device 810A can provide an audio output 1528 corresponding to the voicemail from John (e.g., "Hi, Bill, this is John. Do you have time to have lunch together tomorrow?").

[0283] In some embodiments, the digital assistant operating on device 810A can determine whether the notification is to be provided at device 810A in according with one or more speech inputs. As described above, in some examples, device 810A can be shared among multiple users. Therefore, a notification of an event that device 810A receives may or may be for a particular user (e.g., user 1504). As illustrated in FIG. 15D and continuing the above example where device 810A outputs a response in accordance with the notification of the event (e.g., "Bill, you have a voicemail from John"), device 810A may receive a subsequent speech input 1536 from user 1505, which is a user that is different from the intended user for providing the

notification of the event. Speech input 1536 may include, for example, "I am not Bill. He is not here."

[0284] In some examples, the digital assistant operating on device 810A can obtain an identity of the user who provides one or more speech inputs and determine whether the notification is to be provided to the user who provides the one or more speech inputs. As shown in FIG. 15D, for example, based on speech input 1536 (e.g., "I am not Bill. He is not here"), the digital assistant operating on device 810A can determine that user 1505 is not the user to which the notification is intended to be provided (e.g., "not Bill"). Accordingly, the digital assistant operating on device 810A can determine that the notification should not be provided to user 1505.

[0285] In some examples, to obtain the identity of the user who provides one or more speech inputs, the digital assistant operating on device 810A can obtain authentication data associated with the user. As illustrated in FIG. 15E and continuing the above example where device 810A outputs a response in accordance with the notification of the event (e.g., "Bill, you have a voicemail from John"), the digital assistant operating on device 810A may receive a subsequent speech input 1546 from user 1507, which is a user that is different from the intended user for providing the notification of the event. User 1507 may be, for example, a guest in user 1504's (e.g., Bill) house. User 1507 may decide to listen to Bill's message and thus speech input 1546 may include, for example, "Play the message." In some examples, the digital assistant operating on device 810A can obtain authentication data associated with user 1507. Similar to those described above, the authentication data can include user 1507's voice biometrics, user 1507's facial recognition data, sensing of another device that identifies the user 1507 (e.g., the user's smart watch), and other credentials of the user 1507 (e.g. fingerprints, passwords, etc.). Based on the authentication data, the digital assistant operating on device 810A can obtain a determination of the identity of user 1507. For example, the digital assistant operating on device 810A can authenticate user 1507 based on the authentication data. As another example, the digital assistant operating on device 810A can provide the authentication data to at least one of device 830 (e.g., Bill's smartphone) or device 820 (e.g., a server) for authentication. Device 830 and/or device 820 receive the authentication data and can perform authentication to obtain identity of user 1507 (e.g., matching voice biometrics, fingerprints, passwords, etc.). The digital

assistant operating on device 810A can thus receive user 1507's identity from at least one of device 830 or device 820.

[0286] In some examples, the digital assistant operating on device 810A can determine, based on the identity of the user and based on the received notification, whether the notification should be provided to the user who provides at least one of the one or more speech inputs. For example, as shown in FIG. 15E, based on the identity of user 1507 (e.g., a guest in user 1504's house) and notification 1506 (e.g., a representation of a voicemail for Bill), the digital assistant operating on device 810A may determine that notification 1506 should not be provided to user 1507 because the identity of user 1507 does not match with the user for which the notification is intended (e.g., user 1507 is not Bill or authorized. Accordingly, the digital assistant operating on device 810A can provide an audio output 1548 informing user 1507 that he or she is not authorized to receive the notification (e.g., "Sorry, you are not authorized to listen to this message"). On the other hand, the digital assistant operating on device 810A may determine that notification 1506 should be provided to user 1507 because the identity of user 1507 matches with the user for which the notification is intended (e.g., user 1507 is Bill) or authorized. For example, user 1507 may be user 1504's family member and is authorized to receive notifications for user 1504. Accordingly, device 810A can provide an audio output including the content of the notification 1506.

[0287] In some examples, in accordance with a determination that the notification is to be provided to the user who provides at least one of the one or more speech inputs, the digital assistant operating on device 810A can further determine whether the notification is to be provided at device 810A. As illustrated in FIG. 15F, based on a speech input 1556 (e.g., "Play the message"), device 810A can obtain the identity of user 1504 and determine that user 1504 is authorized to receive the notification. Thus, the digital assistant operation on device 810A can determine that the notification should be provided to user 1504. In some examples, the digital assistant operation on device 810A can further determine whether the notification is to be provided at device 810A. As shown in FIG. 15A, one or more devices may be disposed in the vicinity of user 1504. For example, device 810A may be disposed in living room 1520; device 810B may be disposed in office 1540; and device 810C may be disposed in bedroom 1560. In some examples, device 810A may or may not be the optimum device for providing notifications to user 1504. For example, user 1504 may be moving away from device 810A (e.g., moving

toward office 1540). As another example, there may be other users (e.g., guests) near device 810A disposed in living room 1520, and thus user 1504 may not want to receive notifications from device 810A for privacy concerns.

[0288] Similar to those described above, in some examples, a determination of whether the notification is to be provided at device 810A can be based on the user request represented by the user's speech input (e.g., "Play the message on my office speaker"). In some examples, such determination can be based on at least one of detecting a location of the user or tracking the user's movement. For example, user 1504 may want to go to office 1540 to receive the notification (e.g., listen to a voicemail, pick up a phone call, etc.). The digital assistant operating on device 810A may detect that user 804 is located in living room 1520 but is moving toward office 1540. The digital assistant operating on device 810A can detect location and/or movement using, for example, one or more sensors such as motion sensors, positioning systems, cameras, signal strength measurements to the various devices, etc. Based on the detection of user movement, the digital assistant operating on device 810A can determine that the notification should be provided by device 810B disposed in office 1540, instead of device 810A disposed in living room 1520. In other examples, the digital assistant operating on device 810A may detect that user 804 is not moving and remains in living room 1520. Accordingly, device 810A can determine that the notification should be provided by device 810A disposed in living room 1520.

[0289] FIG. 15G illustrates another example in which device 810A may determine that the notification should be provided by another device. As shown in FIG. 15G, while receiving notification 1572 (e.g., a representation of a voicemail from John), device 810A may be providing audio outputs 1576 (e.g., playing a media item). Thus, the digital assistant operating on device 810A may determine that notification 1572 should not be provided at device 810A to avoid interruption of providing audio output 1576. Accordingly, the digital assistant operating on device 810A can determine an additional device for providing notification 1572. In some examples, such determination is based on context information. For instance, based on information that device 810A is currently providing audio output 1576 and based on detection of device 830, device 810A can determine that the notification 1572 can be provided at device 830. In some examples, device 810A can provide an output (e.g., audio and/or visual output) confirming with user 1504 that notification 1572 should be provided at another device. It is appreciated that the digital assistant operating on device 810A can determine whether the

notification is to be provided at device 810A or at another device based on any context information, such as the user's preferences (e.g., user 1504 prefers to listen voicemail from a colleague on device 810B in office 1540), past devices used for providing notifications, device attributes and capabilities (e.g., device 810B may provide better sound than device 810A), etc.

[0290] In some embodiments, in accordance with a determination that a notification is to be provided at device 810A, the digital assistant operating on device 810A can provide the notification at device 810A. For example, as illustrated in the above examples, digital assistant operating on device 810A can provide audio outputs including the notification (e.g., output voicemails, phone calls, calendar reminders, etc.). In accordance with a determination that a notification is to be provided at a device different from device 810A, the digital assistant operating on device 810A can cause the notification to be provided at the another device. For example, the digital assistant operating on device 810A can forward the notification to device 830, or send a request to device 830 for providing the notification at device 830. Based on the notification or request, device 830 can provide audio outputs 1574 including the content of the notification (e.g., output voicemails, phone calls, calendar reminders, etc.).

6. Process for providing digital assistant services based on user inputs.

[0291] FIGs. 16A-16I illustrates process 1600 for operating a digital assistant for providing digital assistant services based on user inputs, according to various examples. Process 1600 is performed, for example, using one or more electronic devices implementing a digital assistant. In some examples, process 1600 is performed using a client-server system (e.g., system 100), and the blocks of process 1600 are divided up in any manner between the server (e.g., DA server 106) and a client device. In other examples, the blocks of process 1600 are divided up between the server and multiple client devices (e.g., a mobile phone and a smart watch). Thus, while portions of process 1600 are described herein as being performed by particular devices of a client-server system, it will be appreciated that process 1600 is not so limited. In other examples, process 1600 is performed using only a client device (e.g., user device 104, electronic device 810A, device 830, or device 840) or only multiple client devices. In process 1600, some blocks are, optionally, combined, the order of some blocks is, optionally, changed, and some blocks are, optionally, omitted. In some examples, additional steps may be performed in combination with the process 1600.

[0292] With reference to FIG. 16A, at block 1602, a second speech input including a predetermined content (e.g., "Wake up, speaker" or "Hey, Speaker") is received. At block 1604, in response to receiving the second speech input, a first electronic device is activated. The first electronic device can be a service-extension device (e.g., device 810A as shown in FIGs. 8A-15G). At block 1606, in some examples, the second speech input does not cause one or more additional electronic devices to be activated. The one or more additional electronic devices may be disposed in the vicinity of the first electronic device. For example, the volume or sound pressure associated with the speech input can be detected and recorded by both of the first electronic device and an additional electronic device. Based on the comparison of the sound pressure detected at the two devices, the user's position relative to the two devices can be determined. For example, it may be determined that the user is physically closer to the first electronic device than to other device. As a result, the first electronic device may be activated while the other device may not be activated.

[0293] At block 1608, a first speech input representing a user request is received from a first user. At block 1610, the user request comprises a request for information specific to the first user (e.g., the first user's calendar, contacts, etc.). At block 1612, the user request comprises a request for non-user-specific information (e.g., weather information, stock prices, sports game information, etc.) At block 1614, the user request comprises a request for performing a task (e.g., play music, establish a conference, etc.).

[0294] At block 1616, an identity of the first user is the obtained. At block 1618, authentication data associated with the first user is obtained. The authentication data may include, for example, user's voice biometrics, user's facial recognition data, sensing of another electronic device that identifies the user, other credentials of the user, such as the user's fingerprints, passwords, or the like. At block 1620, a determination of the identity of the first user is obtained based on the authentication data. At block 1622, to obtain the identity of the first user, the authentication data are provided to at least one of a second electronic device (e.g., a remote server) or the third electronic device (e.g., user's smartphone). At block 1624, the identity of the first user is received from at least one of the second electronic device or the third electronic device. The identity of the first user is determined based on the authentication data. At block 1626, the identity of the first user is obtained based on the second speech input (e.g.,

"Wake up, speaker" or "Hey, speaker"). As described above, the second speech input may be associated with the user's voice biometrics and can be used for determination of the user's identity.

[0295] With reference to FIG. 16B, at block 1628, one or more speech inquiries regarding the user request represented by the first speech input is outputted. The speech inquiries may be used to clarify the first speech input with the first user (e.g., "What's that?" or "I did not quite get that"). At block 1630, an additional speech input is received from the first user in response to the one or more speech inquiries. For example, the first user may repeat or rephrase the first speech input).

[0296] At block 1632, a connection is established between the first electronic device (e.g., a server-extension device) and at least one of the second electronic device (e.g., a remote server) or the third electronic device (e.g., a device disposed in the vicinity of the first electronic device). At block 1634, establishing a connection is based on a near-field communication between the first electronic device and the third electronic device. At block 1636, establishing a connection is based on detecting of the third electronic device being within a predetermined distance from the first electronic device. At block 1638, establishing a connection is based on a previous established connection between the first electronic device and the third electronic device. For example, the connected can be established based a log file indicating that the first electronic device and the third electronic device have been connected in the past before. The log file may also indicate connection parameters used in the previous connection.

[0297] As described above, a server-extension device can be shared by multiple users and thus connect to multiple devices associated with one or more users. At block 1640, a connection is established between the first electronic device (e.g., a service-extension device) and the third electronic device (e.g., a client device of the first user). The third electronic device is associated with the first user. At block 1642, a connection is established between the first electronic device and a fourth electronic device (e.g., a tablet device of the second user). The fourth electronic device is associated with a second user. At block 1644, in some examples, after establishing a connection between the first electronic device and the third electronic device, the second electronic device is notified of the established connection. For example, after a connection is

established between a server-extension device (e.g., device 810A shown in FIGs. 8A-8B) and a smartphone device, a server can be notified of the established connection.

[0298] With reference to FIG. 16C, as described above, an identity of the first user is obtained. At block 1646, in accordance with the user identity, a representation of the user request is provided to at least one of a second electronic device or a third electronic device. At block 1648, the second electronic device is a server remotely disposed from the first electronic device; and the third electronic device is a client device disposed in the vicinity of the first electronic device. At block 1650, the third electronic device is a proxy device of a server. For example, a client device (e.g., device 840 shown in FIG. 8A) can operate as a proxy device for a server (e.g., device 820 shown in FIG. 8A) to process requests from other devices (e.g., home automation devices such as am intelligent thermostat).

[0299] At block 1652, in some examples, to provide the representation of the user request to at least one of the second electronic device or the third electronic device, it is determined whether the third electronic device (e.g., a client device disposed in the vicinity of the first electronic device) is communicatively coupled to the first electronic device (e.g., a service-extension device). At block 1654, in accordance with a determination that the third electronic device is communicatively coupled to the first electronic device, the representation of the user request is provided to the third electronic device and not to the second electronic device. At block 1656, in accordance with a determination that the third electronic device is not communicatively coupled to the first electronic device, the representation of the user request is provided to the second electronic device.

[0300] At block 1658, in some examples, the representation of the user request is provided to the second electronic device (e.g., a remote server) and not to the third electronic device (e.g., a client device disposed in the vicinity of the first electronic device). At block 1660, in some examples, the representation of the user request is provided to both the second electronic device and the third electronic device.

[0301] As described above, the second electronic device and/or the third electronic device receives the representation of the user request and can determine whether one or both is to provide a response to the first electronic device. With reference to FIG. 16C, at block 1662, based on a determination of whether the second electronic device or the third electronic device,

or both, is to provide a response to the first electronic device, the response to the user request is received from the second electronic device or the third electronic device.

[0302] At block 1664, as described above, in some examples, the representation of the user request is provided to the third electronic device and not to the second electronic device. At block 1666, for receiving the response to the user request, it is caused the third electronic device (e.g., a client device) to determine whether the third electronic device is capable of providing the response to the user request. At block 1668, in accordance with a determination that the third electronic device is capable of providing the response to the user request, it is received, at the first electronic device, the response to the user request from the third electronic device. At block 1670, a determination is made that the third electronic device is incapable of providing the response to the user request. At block 1672, in accordance with such a determination, the representation of the user request is forwarded by the third electronic device to the second electronic device. At block 1674, the response to the user request is received at the first electronic device from the second electronic device.

[0303] With reference to FIG. 16E, as described above, in some examples, at block 1676, the representation of the user request is provided to the second electronic device (e.g., a remote server) and not to the third electronic device (e.g., a client device). At block 1678, for receiving a response to the user request at the first electronic device, it is caused the second electronic device to determine whether the second electronic device is capable of providing the response to the user request. At block 1680, in accordance with a determination that the second electronic device is capable of providing the response to the user request, the response to the user request is received at the first electronic device from the second electronic device. At block 1682, it is determined that the second electronic device is incapable of providing the response to the user request.

[0304] At block 1684, in accordance with such a determination, the representation of the user request is forwarded by the second electronic device to the third electronic device. The third electronic device (e.g., a client device) can thus provide a response based on the user request. At block 1686, the response to the user request.is received at the first electronic device. At block 1688, the first electronic device receives the response to the user request from the third electronic device. At block 1690, the first electronic device receives the response to the user request from

the second electronic device (e.g., a remote server) based on a response provided by the third electronic device to the second electronic device. For example, a client device can forward the response to the remote server, which provides the response to the first electronic device (e.g., a server-extension device).

[0305] With reference to FIG. 16F, as described, in some examples, at block 1692, the representation of the user request is provided from the first electronic device to both the second electronic device and the third electronic device. At block 1694, for receiving the response to the user request, it is caused the second electronic device (e.g., a remote server) to determine whether the second electronic device is capable of providing the response to the user request. At block 1696, for receiving the response to the user request, it is caused the third electronic device (e.g., a client device disposed in the vicinity of the first electronic device) to determine whether the third electronic device is capable of providing the response to the user request. One or both of the determinations in block 1694 and block 1696 can be performed.

[0306] At block 1698, in accordance with a determination that the second electronic device is capable of providing the response to the user request, and that the third electronic device is incapable of providing the response to the user request, the response to the user request is received at the first electronic device from the second electronic device. At block 1700, in accordance with a determination that the third electronic device is capable of providing the response to the user request, and that the second electronic device is incapable of providing the response to the user request, the response to the user request is received at the first electronic device from the third electronic device. At block 1702, in accordance with a determination that both the second electronic device and the third electronic device are capable of providing the response to the user request, the response to the user request is received at the first electronic device from the second electronic device or the third electronic device based on a predetermined condition. The predetermined condition can be, for example, a pre-configured policy (e.g., the third electronic device is the default device to provide a response), user preferences, bandwidth conditions of the connections to the second and third electronic devices, etc.

[0307] With reference to FIG. 16G, at block 1704, it is determining whether the representation of the response is to be provided by the first electronic device (e.g., a service-extension device). At block 1706, the determination of whether the response is to be provided by

the first electronic device is based on the user request (e.g., the user's speech input indicates that the response is to be provided at a different electronic device). At block 1708, the determination of whether the response is to be provided by the first electronic device is based on at least one of detecting a location of the user or tracking the user's movement. For example, if it is detected that the user is moving away from the first electronic device toward another device, the response may not be provided by the first electronic device.

[0308] At block 1710, in accordance with a determination that the representation of the response is to be provided by the first electronic device, the representation of the response is provided to the first user by the first electronic device. At block 1712, in accordance with a determination that the representation of the response is not to be provided by the first electronic device, the response to is forwarded to one or more additional electronic devices, which can provide the response to the first user.

[0309] With reference to FIG. 16H, at block 1714, a representation of the response is provided to the first user. At block 1716, to provide the representation of the response, a speech output including information in response to the user request is provided at the first electronic device. At block 1718, the information is provided by the second electronic device or the third electronic device to the first electronic device.

[0310] At block 1720, to provide the representation of the response, a speech output associated with performing a task in accordance with the user request is provided at the first electronic device. At block 1722, in some examples, the task is performed by the third electronic device (e.g., a client device such as the first user's smartphone). At block 1724, in some examples, the task is performed by the first electronic device and the third electronic device. For example, the first electronic device may output an audio portion of a response, while the third electronic device (e.g., a TV set-top box connected to a TV screen) may output a video portion of the response. At block 1726, the task is further performed by one or more additional electronic devices. For example, in addition to providing a response (e.g., playing a movie) at a service-extension device and a client device such as a TV set-top box, the response may be further provided at an additional device such as a laptop computer.

[0311] At block 1728, in some examples, one or more connections are established between the first electronic device and one or more additional electronic devices. The additional

electronic devices are the same type of devices as the first electronic device. For example, connections can be established among multiple service-extension devices (e.g., devices 810A-C as shown in FIGs. 8A-8B).

[0312] At block 1730, the response is being provided to the first user by the first electronic device. At block 1732, while providing the response to the first user by the first electronic device, it is determined whether the response is to be continually provided at a different electronic device (e.g., a client device such as the user's smartphone). At block 1734, the determination of whether the response is to be continually provided at a different electronic device is based on a third speech input (e.g., a speech input from the first user such as "Continue to play the song on my phone"). At block 1736, the determination of whether the response is to be continually provided at a different electronic device is based on detecting whether the first user's location variation with respect to the first electronic device satisfies a predetermined condition. For example, it can be determined whether the first user has moved out of a predetermined boundary such that the response should be continually provided at a device different from the first electronic device.

[0313] At block 1738, in accordance with a determination that the response is to be continually provided at a different electronic device, the response is caused to be continually provided by at least one of the third electronic device or one or more additional electronic devices.

[0314] At block 1740, after providing the response to the first user, subsequent speech inputs are monitored.

7. Process for providing digital assistant services based on notifications of events.

[0315] FIGs. 17A-17D illustrates process 1800 for operating a digital assistant for providing digital assistant services based on notifications of events, according to various examples. Process 1800 is performed, for example, using one or more electronic devices implementing a digital assistant. In some examples, process 1800 is performed using a client-server system (e.g., system 100), and the blocks of process 1800 are divided up in any manner between the server (e.g., DA server 106) and a client device. In other examples, the blocks of process 1800 are divided up between the server and multiple client devices (e.g., a mobile phone and a smart

watch). Thus, while portions of process 1800 are described herein as being performed by particular devices of a client-server system, it will be appreciated that process 1800 is not so limited. In other examples, process 1800 is performed using only a client device (e.g., user device 104, devices 810A-C) or only multiple client devices. In process 1800, some blocks are, optionally, combined, the order of some blocks is, optionally, changed, and some blocks are, optionally, omitted. In some examples, additional steps may be performed in combination with the process 1800.

[0316] With reference to FIG. 17A, at block 1802, prior to receiving a notification of an event, a connection is established between a first electronic device (e.g., a service-extension device) and at least one of a second electronic device (e.g., a server) or a third electronic device (e.g., a client device disposed in the vicinity of the first electronic device). At block 1804, establishing the connection is based on a near-field communication between the first electronic device and the third electronic device. At block 1806, establishing the connection is based on detecting of the third electronic device being within a predetermined distance from the first electronic device. At block 1808, establishing the connection is based on a previous established connection between the first electronic device and the third electronic device.

[0317] As described above, the first electronic device (e.g., a service-extension device) can be shared by multiple users. At block 1810, a connection between the first electronic device and the third electronic device is established; and the third electronic device is associated with the first user. At block 1812, a connection between the first electronic device and a fourth electronic device is established; and the fourth electronic device is associated with a second user. At block 1814, after establishing a connection between the first electronic device and the third electronic device, the second electronic device (e.g., a remote server) is notified of the established connection.

[0318] With reference to FIG. 17B, at block 1816, a notification of an event associated with a first user is received. At block 1818, the notification of the event includes a representation of at least one of an incoming call, a reminder, a message, a voicemail, or a news alert. At block 1820, the notification is received from at least one of a second electronic device (e.g., a remote

server) or a third electronic device (e.g., a client device disposed in the vicinity of the first electronic device).

[0319] At block 1822, in response to receiving the notification, an indication of the notification is outputted. An indication can be, for example, a beep, an alert, a ringtone, etc. At block 1824, the indication of the notification is outputted by the first electronic device or one of the additional electronic devices communicatively coupled to the first electronic device. For example, a client device such as user's smartphone can output the indication of the notification; and another service-extension device can output the indication of the notification.

[0320] At block 1826, one or more speech inputs are received. At block 1828, for example, a first speech input is received regarding the notification (e.g., the user may provide a first speech input inquiring about the indication of the even notification such as "What is it?"). At block 1830, a response is outputted in accordance with the notification of the event. For example, a speech output may be provided such as "You have a voicemail from John." At block 1832, a second speech input is received. For example, the user may say "Play the voicemail."

[0321] With reference to FIG. 17C, at block 1834, in accordance with the one or more speech inputs, it is determined whether the notification is to be provided at the first electronic device. At block 1836, to determine whether the notification is to be provided at the first electronic device, it is obtained an identity of the user who provides at least one of the one or more speech inputs. At block 1838, to obtain the identity of the user, it is obtained authentication data associated with the user who provides at least one of the one or more speech inputs. The authentication data can include, for example, the user's biometrics, fingerprints, facial recognition data, passwords, etc. At block 1840, it is obtained a determination of the identity of the user who provides at least one of the one or more speech inputs based on the authentication data.

[0322] At block 1842, to obtain the determination of the identity of the user who provides at least one of the one or more speech inputs, the authentication data is provided to at least one of a second electronic device and a third electronic device. At block 1844, it is received the identity of the user who provides at least one of the one or more speech inputs from at least one of the

second electronic device and the third electronic device. The identity of the user who provides at least one of the one or more speech inputs is determined based on the authentication data.

[0323] At block 1846, it is determined, based on the identity of the user who provides at least one of the one or more speech inputs and based on the notification, whether the notification is to be provided to the user who provides at least one of the one or more speech inputs. At block 1848, in accordance with a determination that the notification is to be provided to the user who provides at least one of the one or more speech inputs, it is determined whether the notification is to be provided at the first electronic device.

[0324] With reference to FIG. 17D, at block 1850, in accordance with a determination that the notification is to be provided at the first electronic device, the notification is provided at the first electronic device. At block 1852, to provide the notification at the first electronic device, an audio output is provided associated with the notification at the first electronic device.

[0325] At block 1854, in accordance with a determination that the notification is not to be provided at the first electronic device, an additional electronic device for providing the notification is determined. At block 1856, determining the additional electronic device for providing the notification is based on the one or more speech inputs. At block 1858, determining the additional electronic device for providing the notification is based on context information.

[0326] The operations described above with reference to FIGs. 16A-16I and 17A-17D are optionally implemented by components depicted in FIGs. 1-4, 6A-B, and 7A-C. For example, the operations of process 1600 and 1800 may be implemented by digital assistant system 700. It would be clear to a person having ordinary skill in the art how other processes are implemented based on the components depicted in FIGS. 1-4, 6A-B, and 7A-C.

8. Exemplary functions of providing digital assistant services using multiple devices.

[0327] As described above, digital assistant services can be provided by one or more devices. Due to device capability limitations, certain devices may be incapable of, or not optimum for, providing certain digital assistant services. For example, a smartwatch typically has a small screen size and thus is not optimum for playing video. As another example, unlike a smartphone, a TV set-top box may be incapable of providing speech outputs for text messages.

[0328] FIGs. 18A-18E illustrate functionalities for providing digital assistant services based on capabilities of multiple electronic devices, according to various examples. In some examples, the digital assistant (e.g., digital assistant system 700) is implemented by a user device according to various examples. In some examples, the user device, a server (e.g., server 108, device 820), or a combination thereof, may implement a digital assistant system (e.g., digital assistant system 700). The user device can be implemented using, for example, device 200, 400, 600, 820, 830, 840, 1880, and/or 1882. In some examples, the user device is a device having audio outputting capabilities and network connectivity, a smart watch, smartphone, a laptop computer, a desktop computer, or a tablet computer.

[0329] As illustrated in FIG. 18A, user 804 may provide a speech input 1886 such as "Show me the video I took last Sunday." Speech input 1886 can represent a user request (e.g., a request for information or a request for performing a task). In some examples, the digital assistant operating on device 1880 receives speech input 1886 from user 804. Device 1880 can be, for example, a client device (e.g., a wearable device such as smart watch). Device 1880 can also be a device similar to device 810A described above, which can include one or more audio input and output devices (e.g., a microphone and one or more speakers) and one or more network communication interfaces. Device 1880 may or may not be capable of, or optimum for, responding to the user request (e.g., providing the requested information or performing the requested tasks). For example, device 1880 may not have a display or may have a small-sized display that is not optimum to play video.

[0330] In some embodiments, the digital assistant operating on device 1880 can obtain capability data associated with one or more electronic devices capable of being communicatively coupled to device 1880. For example, as shown in FIG. 18A, the digital assistant operating on device 1880 can determine that device 820 (e.g., a remote server), device 830 (e.g., a client device such as the user's smartphone), device 840 (a TV set-top box) are communicatively coupled to device 1880. The determination can be made, for example, via Bluetooth pairing, Wi-Fi connection, etc., similar to as described above with respect to device 810A. Based on the determination that devices 820, 830, and 840 are communicatively coupled, the digital assistant operating on device 1880 can obtain capability data associated with these devices. In some examples, some devices are client devices disposed in the vicinity of device 1880 and some

devices are disposed remotely from device 1880. For example, device 1880, device 830, and device 840 are client devices disposed within a predetermined boundary (e.g., a house, a building, a car, etc.); and device 820 is a server disposed remotely.

[0331] In some examples, capability data can include device capabilities associated with electronic devices capable of being communicatively coupled to device 1880. Device capabilities can include one or more physical capabilities and/or informational capabilities. Physical capabilities of a device can include the device's physical attributes such as whether the device has a display, the size of the display, the number of speakers, network capabilities, or the like. Informational capabilities of a device can include data that the device is capable of providing. For example, device 830 may store media items (e.g., videos and photos) that user 804 took, and is thus capable of providing the stored media items to other devices communicatively connected to device 830.

[0332] In some examples, prior to obtaining the capability data, the digital assistant operating on device 1880 can be configured to establish communication with other electronic devices (e.g., devices 820, 830, and/or 840). In some examples, the communication can be established via a direct communication connection, such as Bluetooth, near-field communication (NFC), BTLE (Bluetooth Low Energy), or the like, or via a wired or wireless network, such as a local Wi-Fi network. For example, the digital assistant operating on device 1880 can detect device 830 via Bluetooth discovery, and communicatively coupled to devices 830 via Bluetooth connection. As another example, the digital assistant operating on device 1880 can detect a Wi-Fi network and couple to device 840 via the Wi-Fi network. As another example, the digital assistant operating on device 1880 can detect a near field communication when the device 830 (e.g., a client device such as the user's smartphone) is in close proximity with, or physically in touch with, device 1880. For instance, to pair up device 1880 and device 830, user 804 may tap device 1880 with device 830, thereby establishing near-field communication between the two devices. As another example, the digital assistant operating on device 1880 may detect that device 830 (e.g., a client device such as the user's smartphone) is within a predetermined distance (e.g., within a range of Bluetooth communication) and establish a connection with device 830. For instance, as user 804 approaches device 1880 with device 830, the digital assistant operating on device 1880 detects that device 830 is within communication range, and thus connects device 1880 with device 830.

As another example, the digital assistant operating on device 1880 may establish a connection with device 830 based on one or more previous established connections between the two devices. For instance, the digital assistant operating on device 1880 can store a log file including the devices that it connected in the past, and optionally connection parameters. Thus, based on the log file, the digital assistant operating on device 1880 can determine, for example, that it has connected to device 830 before. Based on such determination, the digital assistant operating on device 1880 can establish the connection with device 830 again.

[0333] In some examples, prior to obtaining the capability data, the digital assistant operating on device 1880 can inquire user 804 regarding accessing one or more devices capable of being communicatively coupled to device 1880. For example, the digital assistant operating on device 1880 may provide a speech output such as "Do I have your permission to access your phone and TV?" In some examples, user 804 may respond with a speech input either permitting the access or denying the access. In response to receiving the speech input, the digital assistant operating on device 1880 can determine whether it is authorized to access the devices communicatively coupled to device 1880. For example, if user 804's speech input is "OK," the digital assistant operating on device 1880 can determine that it is authorized to access devices 830 and 840. If user 804's speech input is "No," the digital assistant operating on device 1880 can determine that it is not authorized to access devices 830 and 840. If user 804's speech input is "Yes for my phone, No for my TV" (e.g., user 804 may be watching another video by using device 840 and device 1882 (e.g., a TV display), and does not wish to disturb the video playing on device 1882), the digital assistant operating on device 1880 can determine that it is authorized to access device 830 but not device 840 for playing the video on device 1882.

[0334] With reference to FIGs. 18A and 18B, in some embodiments, in accordance with the capability data, the digital assistant operating on device 1880 can identify, from the one or more electronic devices capable of being communicatively coupled to the device, a device for providing at least a portion of a response to the user request. In some examples, the digital assistant operating on device 1880 can obtain one or more steps for responding to the user request based on speech input 1886. The one or more steps for responding to the user request can include steps for providing requested information and/or steps for performing a requested task. For example, based on speech input 1886 such as "Show me the video I took last Sunday,"

device 1880 can determine the user request is to find the video user 804 took last Sunday and play it. The determination can be made using natural language processing techniques described above.

[0335] According to the determined user request, the digital assistant operating on device 1880 can determine one or more steps required for responding to the user request. For example, the digital assistant operating on device 1880 can determine that step # 1 for playing a video user 804 took last Sunday is to find the particular video that user 804 took last Sunday; and step # 2 is to play the particular video. In some embodiments, the determination of the one or more steps can be made on another device (e.g., device 820 such as a remote server) and provided to device 1880. In some embodiments, the determination of the one or more steps can be made using both device 820 and device 1880. For example, a digital assistant can operate on device 1880 in the front end to interface with user 804 and operate on device 820 in the back end to process the user input. In some embodiments, the one or more steps for responding to the user request can form at least a portion of an execution plan. An execution plan may include the steps for responding to the user request and the device for perform each of the steps.

[0336] In some embodiments, the digital assistant operating on device 1880 can identify one or more devices for performing the steps for responding to the user request based on capability data associated with one or more electronic devices capable of being communicatively coupled to device 1880. Continuing the above example, the digital assistant operating on device 1880 can identify a device for performing step # 1 of finding a particular video that user 804 took last Sunday; and identify a device for performing step # 2 of playing the particular video. As shown in FIG. 18B, the digital assistant operating on device 1880 may determine that, among the devices communicatively connected to device 1880 (e.g., devices 820, 830, and 840), the capability data of device 830 (e.g., a client device such as the user's smartphone) indicate that device 830 has the capability of finding the particular video user 804 took last Sunday. For example, user 804 took a video last Sunday using device 830, and therefore the informational capability data of device 830 may indicate that a file stored in device 830 has a format of a video and a time stamp of last Sunday. Accordingly, the digital assistant operating on device 1880 can identify device 830 for performing step # 1 of finding the particular video user 804 intended.

[0337] As another example, the digital assistant operating on device 1880 may determine that, among devices 820, 830, and 840, the capability data of device 840 (e.g., a TV set-top box) indicate that device 840 is an optimum device for performing step # 2 of playing videos. For example, device capability data of devices 830 and 840 may both indicate that the devices are capable of playing videos. Device capability data of device 840 may further indicate that one or more device attributes (e.g., display size/resolution/number of speakers) of a device 1882 (e.g., a TV screen), on which device 840 can play the video, are superior than the device attributes of device 830. For example, the display size of device 1882 is bigger than the display size of device 830. As a result, the digital assistant operating on device 1880 can identify device 840, instead of device 830, for performing step # 2 of playing the video.

[0338] In some embodiments, as shown in FIG. 18B, based on the determination of the devices for performing the steps for responding to the user request, the digital assistant operating on device 1880 can provide a speech output to user 804 seeking confirmation or permission to access the devices identified for responding to the user request. For example, as shown in FIG. 18B, the digital assistant operating on device 1880 can provide a speech output 1883 such as "I will retrieve the video from your phone and play it on your TV, OK to proceed?" With reference to FIG. 18C, in some examples, device 1880 may receive subsequent speech input 1885 from user 804. In response to receiving speech input 1885, the digital assistant operating on device 1880 can determine whether at least a portion of the response is to be provided by one or more devices communicatively coupled to device 1880. For example, if speech input 1885 includes "OK," the digital assistant operating on device 1880 can determine that the video should be retrieved from device 830 and played on device 1882 (e.g., a TV display) using device 840 (e.g., a TV set-top box). If speech input 1885 includes "No, play the video on my phone," the digital assistant operating on device 1880 can determine that the video should be retrieved from device 830, but played on device 830 rather than device 1882. It is appreciated that, in some examples, providing speech output 1883 seeking confirmation/permission and receiving subsequence user input 1885 are optional.

[0339] In some embodiments, prior to providing a speech output for confirming or requesting permission to user the devices identified for responding to the user request, the digital assistant can annotate one or more steps for responding to the user request, and providing the

speech output based on the annotation. Using the above example, the digital assistant operation on device 1880 can determine whether performing a particular step would require altering a state of a device. For instance, for performing step # 2 of playing video on device 1882 (e.g., a TV display), a state of device 840 may be altered (e.g., changing from a state of power off to power on, switching from the current playing video to the video user requested, etc.). As a result, the digital assistant operating on device 1880 can annotate step # 2 as a step that requires altering state of the identified device.

[0340] As another example, for performing step # 1 of finding the video user 804 took on last Sunday, the digital assistant operating on device 1880 may determine that performing step # 1 would not require altering the state of device 830. As a result, the digital assistant operating on device 1880 may not annotate step # 1 as a step that requires altering state of the identified device. In some examples, based on the annotation, the digital assistant operating on device 1880 can then provide a speech output seeking confirmation or permission to use the devices identified for responding to the user request. Using the above example, because step # 2 of playing video is a state-altering step for the identified device 840, the digital assistant operating on device 1880 can provide a speech output seeking permission to access device 840. The speech output may include, for example, "I will play the video on your TV, OK to proceed?" And because step # 1 of finding a video is a not state-altering step for the identified device 830, the digital assistant operating on device 1880 may not provide a speech output seeking permission to access device 830.

[0341] With reference to FIG. 18C, in some embodiments, the digital assistant operating on device 1880 can cause one or more identified devices to provide at least a portion of a response to the user request. For example, the digital assistant operating on device 1880 can request device 830 to search and find the video user 804 took last Sunday and transmit the video to device 840. It can further request device 840 to play the video on device 1882 (e.g., a TV).

[0342] As described above, in some examples, prior to obtaining the capability data, the digital assistant operating on device 1880 can seek confirmation or permission to access one or more electronic devices being capable of communicatively coupled to device 1880. For example, the digital assistant operating on device 1880 may provide a speech output such as "Do

I have your permission to access your phone?" With reference to FIG. 18D, in some embodiments, the digital assistant operating on device 1880 can provide one or more duration options for accessing the devices being capable of communicatively coupled to device 1880. For example, as illustrated in FIG. 18D, the digital assistant operating on device 1880 can display options 1884A-C on device 1880. Option 1884A may include "Allow Once," indicating that the access to device 830 from device 1880 is permitted only this time. Option 1884B may include "Allow while both devices are at home," indicating that the access to device 830 from device 1880 is permitted while both devices are within a predetermined boundary (e.g., within or nearby a house). Option 1884C may include "Always allow," indicating the access to device 830 from device 1880 is always permitted. In some examples, the digital assistant may also provide an option 1884D (not shown), which may include "Not allow," indicating that access to device 830 from device 1880 is denied. In some embodiments, similar duration options 1887A-D can be displayed on device 830, thereby enabling the user of device 830 (e.g. a user that may or may not be the same as the user of device 1880) to control the access of device 830.

[0343] In some embodiments, the digital assistant operating on device 1880 can receive a selection of a duration option from user 804 and access the devices capable of being communicatively coupled to device 1880 based on the selected duration option. For example, if the selection is option 1884A such as "Allow once," the digital assistant operating on device 1880 may only access device 830 to find the video the user requested for this time.

[0344] As described above, in some embodiments, a digital assistant can cause one or more identified devices to provide at least a portion of a response to the user request. In some embodiments, prior to causing the identified devices to provide a response, the digital assistant can obtain an identity of the user and determine whether the user is authorized to receive at least a portion of the response. As illustrated in FIG. 18E, for example, device 840 may be a device (e.g., a TV set-top box) that is shared between multiple users and thus user 1888 may be a user that is authorized to access device 840. Device 830 may be a client device such as a smartphone of another user (e.g., user 804). User 1888 may not be authorized to access device 830. In some examples, user 1888 may provide a speech input 1889 such as "Play the video Bill Took last Sunday on his phone." The digital assistant operating on device 840 may identify, based on the capability data of devices communicatively coupled to device 840, device 830 for providing at

least a portion of a response to the user request. In some embodiments, before accessing device 830 to perform a step for responding to the user request, the digital assistant operating on device 840 can obtain the identity of user 1888. In some examples, obtaining the identity of user 1888 can be based on a voice profile. A voice profile may include voice biometrics, such as the user's voice characteristics (e.g., acoustic patterns, voiceprints, the user's accent, or the like). A voice profile can be associated with a particular user and uniquely identifies the user. For example, a voice profile of user 1888 can include voice characteristics of user 1888 and thus uniquely identify user 1888. In some examples, a voice profile can also assistant the natural language processing described above to more accurately determine the user intent. For example, the speech-to-text conversion process may be more accurately performed using a voice profile that includes the user's accent data.

[0345] With reference to FIG. 18E, the digital assistant operating on device 840 can compare the voice characteristics in speech input 1889 with one or more voice profiles of one or more authorized users of device 830. Based on the comparison, the digital assistant operating on device 840 may determine that voice characteristics in speech input 1889 does not match any of the voice profile for authorized user of device 830. As a result, the digital assistant operating on device 840 can determine that user 1888 is not authorized to access device 830, and thus not authorized to access the video stored in device 830.

9. Process for providing digital assistant services using multiple devices.

[0346] FIGs. 19A-19D illustrates process 1900 for operating a digital assistant for providing digital assistant services based on notifications of events, according to various examples. Process 1900 is performed, for example, using one or more electronic devices implementing a digital assistant. In some examples, process 1900 is performed using a client-server system (e.g., system 100), and the blocks of process 1900 are divided up in any manner between the server (e.g., DA server 106) and a client device. In other examples, the blocks of process 1900 are divided up between the server and multiple client devices (e.g., a mobile phone and a smart watch). Thus, while portions of process 1900 are described herein as being performed by particular devices of a client-server system, it will be appreciated that process 1900 is not so limited. In other examples, process 1900 is performed using only a client device (e.g., user

device 104, device 1880) or only multiple client devices. In process 1900, some blocks are, optionally, combined, the order of some blocks is, optionally, changed, and some blocks are, optionally, omitted. In some examples, additional steps may be performed in combination with the process 1900.

[0347] With reference to FIG. 19A, at block 1902, a first speech input representing a user request is received from a first user. At block 1904, prior to obtaining capability data associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device, a connection is established between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device. In some examples, the first electronic device and the electronic devices capable of being communicatively coupled to the first electronic device are disposed within a predetermined boundary (e.g., a house). In some examples, establishing the connection is based on a near-field communication between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device. In some examples, establishing the connection is based on detecting of the one or more electronic devices capable of being communicatively coupled to the first electronic device being within a predetermined distance from the first electronic device. In some examples, establishing the connection is based on one or more previous established connections between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

[0348] At block 1906, prior to obtaining capability data associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device, the first user is inquired regarding accessing, by the first electronic device, the one or more electronic devices capable of being communicatively coupled to the first electronic device. At block 1908, a third speech input is received from the first user. The third speech input may indicate whether the first electronic device is authorized to access other devices. At block 1910, in response to receiving the third speech input, it is determined whether the first electronic device is authorized to access the one or more electronic devices capable of being communicatively coupled to the first electronic device.

[0349] At block 1912, it is provided one or more duration options for accessing the one or more electronic devices capable of being communicatively coupled to the first electronic device. The duration options may include, for example, allow once, allow while both devices are at home, always allow, not allow. At block 1914, a selection of a duration option is received from the first user. At block 1916, the one or more electronic devices capable of being communicatively coupled to the first electronic device are accessed based on the selected duration option.

[0350] With reference to FIG. 19B, at block 1918, it is obtained capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device. At block 1920, to obtain capability data, it is obtained device capabilities associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device. At block 1922, the device capabilities include one or more physical attributes associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device. At block 1924, the device capabilities include data that are capable of being provided by the one of more electronic devices capable of being communicatively coupled to the first electronic device.

[0351] At block 1926, in accordance with the capability data, it is identified, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request. At block 1928, to identify the second electronic device, one or more steps for responding to the user request are obtained based on the first speech input. In some examples, at block 1930, to obtain the one or more steps, a plan for responding to the user request is received from a third electronic device (e.g., a server) remotely located from the first electronic device. In some examples, at block 1932, a plan for responding to the user request is determined by the first electronic device (e.g., a client device such as a wearable device), wherein the plan comprising one or more steps for responding to the user request.

[0352] At block 1934, it is identified, based on the capability data, the second electronic device (e.g., device 840 such as a TV set-top box) for performing at least one step for responding

to the user request. At block 1936, it is identified, based on the capability data, one or more additional electronic devices for performing the remaining steps for responding the user request.

[0353] With reference to FIG. 19C, at block 1938, a first speech output is provided to the first user regarding providing at least a portion of the response by the second electronic device. For example, the first speech output may be a speech out requesting authorization to access the second electronic device. At block 1940, for providing the first speech output, one or more steps for responding to the user request are annotated. For example, some steps may be annotated as state-altering steps and thus may require authorization; and some steps may not be annotated and thus may not require authorization. At block 1942, the first speech output is provided to the first user based on the annotation of one or more steps.

[0354] At block 1944, a second speech input is received from the first user. The second speech input may indicate whether the first user authorizes accessing certain devices. At block 1946, in response to receiving the second speech input, it is determined whether at least a portion of the response is to be provided by the second electronic device.

[0355] At block 1948, in some examples, prior to causing the second electronic device to provide at least a portion of the response to the first user, an identity of the first user is obtained. At block 1950, the identity is obtained based on a voice profile. At block 1952, it is determined, based on the identity of the first user, whether the first user is authorized to receive at least a portion of the response to the user request.

[0356] With reference to FIG. 19D, at block 1954, the second electronic device is caused to provide at least a portion of the response to the user request. At block 1956, the second electronic device is caused to perform at least one step for responding to the user request. At block 1958, one or more additional electronic devices are caused to perform the remaining steps for responding to the user request.

[0357] The operations described above with reference to FIGs. 19A-19D are optionally implemented by components depicted in FIGs. 1-4, 6A-B, and 7A-C. For example, the operations of process 1900 may be implemented by digital assistant system 700. It would be

clear to a person having ordinary skill in the art how other processes are implemented based on the components depicted in FIGS. 1-4, 6A-B, and 7A-C.

[0358] In accordance with some implementations, a computer-readable storage medium (e.g., a non-transitory computer readable storage medium) is provided, the computer-readable storage medium storing one or more programs for execution by one or more processors of an electronic device, the one or more programs including instructions for performing any of the methods or processes described herein.

[0359] In accordance with some implementations, an electronic device (e.g., a portable electronic device) is provided that comprises means for performing any of the methods or processes described herein.

[0360] In accordance with some implementations, an electronic device (e.g., a portable electronic device) is provided that comprises a processing unit configured to perform any of the methods or processes described herein.

[0361] In accordance with some implementations, an electronic device (e.g., a portable electronic device) is provided that comprises one or more processors and memory storing one or more programs for execution by the one or more processors, the one or more programs including instructions for performing any of the methods or processes described herein.

[0362] Exemplary methods, non-transitory computer-readable storage media, systems and electronic devices are set out in the following items:

1. A method for providing a digital assistant service, comprising:
at a first electronic device with one or more processors and memory:
receiving, from a first user, a first speech input representing a user request;
obtaining an identity of the first user;
in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device;
receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

providing a representation of the response to the first user.

2. The method of item 1, wherein the second electronic device is a server remotely disposed from the first electronic device; and the third electronic device is a client device disposed in the vicinity of the first electronic device.

3. The method of any of items 1 and 2, wherein the third electronic device is a proxy device of a server.

4. The method of any of items 1-3, wherein the user request comprises a request for information specific to the first user.

5. The method of any of items 1-4, wherein the user request comprises a request for non-user-specific information.

6. The method of any of items 1-5, wherein the user request comprises a request for performing a task.

7. The method of any of items 1-6, wherein obtaining the identity of the first user comprises:

obtaining authentication data associated with the first user; and

obtaining a determination of the identity of the first user based on the authentication data.

8. The method of item 7, wherein obtaining the determination of the identity of the first user based on the authentication data comprises:

providing the authentication data to at least one of the second electronic device or the third electronic device; and

receiving the identity of the first user from at least one of the second electronic device or the third electronic device, wherein the identity of the first user is determined based on the authentication data.

9. The method of any of items 1-8, wherein providing the representation of the user request to at least one of the second electronic device or the third electronic device comprises:

determining whether the third electronic device is communicatively coupled to the first electronic device;

in accordance with a determination that the third electronic device is communicatively coupled to the first electronic device, providing the representation of the user request to the third electronic device and not to the second electronic device; and

in accordance with a determination that the third electronic device is not communicatively coupled to the first electronic device, providing the representation of the user request to the second electronic device.

10. The method of any of items 1-9, wherein providing the representation of the user request to at least one of the second electronic device or the third electronic device comprises:

providing the representation of the user request to the second electronic device and not to the third electronic device.

11. The method of any of items 1-8, wherein providing the representation of the user request to at least one of the second electronic device or the third electronic device comprises:

providing the representation of the user request to both the second electronic device and the third electronic device.

12. The method of any of items 1-9, wherein the representation of the user request is provided to the third electronic device and not to the second electronic device; and wherein receiving the response to the user request comprises:

causing the third electronic device to determine whether the third electronic device is capable of providing the response to the user request; and

in accordance with a determination that the third electronic device is capable of providing the response to the user request, receiving, at the first electronic device, the response to the user request from the third electronic device.

13. The method of item 12, further comprising:

in accordance with a determination that the third electronic device is incapable of providing the response to the user request,

forwarding, by the third electronic device, the representation of the user request to the second electronic device; and

receiving, at the first electronic device, the response to the user request from the second electronic device.

14. The method of any of items 1-10, wherein the representation of the user request is provided to the second electronic device and not to the third electronic device, and wherein receiving the response to the user request comprises:

causing the second electronic device to determine whether the second electronic device is capable of providing the response to the user request; and

in accordance with a determination that the second electronic device is capable of providing the response to the user request, receiving, at the first electronic device, the response to the user request from the second electronic device.

15. The method of item 14, further comprising:

in accordance with a determination that the second electronic device is incapable of providing the response to the user request,

forwarding, by the second electronic device, the representation of the user request to the third electronic device; and

receiving, at the first electronic device, the response to the user request.

16. The method of item 15, wherein the first electronic device receives the response to the user request from the third electronic device.

17. The method of item 15, wherein the first electronic device receives the response to the user request from the second electronic device based on a response provided by the third electronic device to the second electronic device.

18. The method of any of items 1-8 and 11, wherein the representation of the user request is provided from the first electronic device to both the second electronic device and the third electronic device, and wherein receiving the response to the user request comprises at least one of:

causing the second electronic device to determine whether the second electronic device is capable of providing the response to the user request; and

causing the third electronic device to determine whether the third electronic device is capable of providing the response to the user request.

19. The method of item 18, further comprising:

in accordance with a determination that the second electronic device is capable of providing the response to the user request, and that the third electronic device is incapable of providing the response to the user request,

receiving, at the first electronic device, the response to the user request from the second electronic device.

20. The method of item 18, further comprising:

in accordance with a determination that the third electronic device is capable of providing the response to the user request, and that the second electronic device is incapable of providing the response to the user request,

receiving, at the first electronic device, the response to the user request from the third electronic device.

21. The method of item 18, further comprising:

in accordance with a determination that both the second electronic device and the third electronic device are capable of providing the response to the user request,

receiving, at the first electronic device, the response to the user request from the second electronic device or the third electronic device based on a predetermined condition.

22. The method of any of items 1-21, wherein providing the representation of the response to the first user comprises:

providing, at the first electronic device, a speech output including information in response to the user request.

23. The method of item 22, wherein the information is provided by the second electronic device or the third electronic device.

24. The method of any of items 1-23, wherein providing the representation of the response to the first user comprises:

providing, at the first electronic device, a speech output associated with performing a task in accordance with the user request.

25. The method of item 24, wherein the task is performed by the third electronic device.

26. The method of any of items 24 and 25, wherein the task is performed by the first electronic device and the third electronic device.

27. The method of item 26, wherein the task is further performed by one or more additional electronic devices.

28. The method of any of items 1-27, further comprising, prior to providing the representation of the response to the first user:

determining whether the representation of the response is to be provided by the first electronic device;

in accordance with a determination that the representation of the response is to be provided by the first electronic device, providing the representation of the response to the first user by the first electronic device; and

in accordance with a determination that the representation of the response is not to be provided by the first electronic device, forwarding the response to one or more additional electronic devices.

29. The method of item 28, wherein determining whether the response is to be provided by the first electronic device is based on the user request.

30. The method of any of items 28 and 29, wherein determining whether the response is to be provided by the first electronic device is based on at least one of detecting a location of the user or tracking the user's movement.

31. The method of any of items 1-30, further comprising, prior to receiving the first speech input,

receiving a second speech input including a predetermined content; and

in response to receiving the second speech input, activating the first electronic device.

32. The method of item 31, wherein the second speech input does not cause one or more additional electronic devices to be activated, the one or more additional electronic devices being disposed in the vicinity of the first electronic device.

33. The method of any of items 31 and 32, wherein obtaining the identity of the first user is based on the second speech input.

34. The method of any of items 1-33, further comprising, prior to providing the representation of the user request to at least one of the second electronic device and the third electronic device:

outputting one or more speech inquiries regarding the user request; and

receiving an additional speech input from the first user in response to the one or more speech inquiries.

35. The method of any of items 1-34, further comprising, prior to providing the representation of the user request to at least one of the second electronic device or the third electronic device:

establishing a connection between the first electronic device and at least one of the second electronic device or the third electronic device.

36. The method of item 35, wherein establishing the connection is based on a near-field communication between the first electronic device and the third electronic device.

37. The method of any of items 35 and 36, wherein establishing the connection is based on detecting of the third electronic device being within a predetermined distance from the first electronic device.

38. The method of any of items 35-37, wherein establishing the connection is based on a previous established connection between the first electronic device and the third electronic device.

39. The method of any of items 35-38, further comprising:

establishing a connection between the first electronic device and the third electronic device, wherein the third electronic device is associated with the first user; and

establishing a connection between the first electronic device and a fourth electronic device, wherein the fourth electronic device is associated with a second user.

40. The method of any of items 35-39, further comprising:

after establishing a connection between the first electronic device and the third electronic device, notifying the second electronic device of the established connection..

41. The method of any of items 1-39, further comprising, establishing one or more connections between the first electronic device and one or more additional electronic devices, wherein the additional electronic devices are the same type of devices as the first electronic device.

42. The method of any of items 1-41, further comprising, while providing the response to the first user by the first electronic device,

determining whether the response is to be continually provided at a different electronic device; and

in accordance with a determination that the response is to be continually provided at a different electronic device, causing the response to be continually provided by at least one of the third electronic device or one or more additional electronic devices.

43. The method of item 42, wherein the determination of whether the response is to be continually provided at a different electronic device is based on a third speech input.

44. The method of any of items 42 and 43, wherein the determination of whether the response is to be continually provided at a different electronic device is based on detecting whether the first user's location variation with respect to the first electronic device satisfies a predetermined condition.

45. The method of any of items 1-44, further comprising, after providing the response to the first user, monitoring subsequent speech inputs.

46. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors

of a first electronic device, cause the first electronic device to:

receive, from a first user, a first speech input representing a user request;

obtain an identity of the first user;

in accordance with the user identity, provide a representation of the user request to at least one of a second electronic device or a third electronic device;

receive, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

provide a representation of the response to the first user.

47. A first electronic device, comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for:

receiving, from a first user, a first speech input representing a user request;

obtaining an identity of the first user;

in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device;

receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

providing a representation of the response to the first user.

48. A first electronic device, comprising:

means for receiving, from a first user, a first speech input representing a user request;

means for obtaining an identity of the first user;

in accordance with the user identity, means for providing a representation of the user request to at least one of a second electronic device or a third electronic device;

means for receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

means for providing a representation of the response to the first user.

49. A first electronic device comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for performing the method of any one of items 1-45.

50. A first electronic device, comprising:

means for performing any of the methods of items 1-45.

51. A non-transitory computer-readable storage medium comprising one or more programs for execution by one or more processors of a first electronic device, the one or more programs including instructions which, when executed by the one or more processors, cause the first electronic device to perform the method of any one of items 1-45.

52. A system for operating a digital assistant, the system comprising means for performing any of the methods of items 1-45.

53. A method for providing a digital assistant service, comprising:

at a first electronic device with one or more processors and memory:

receiving a notification of an event associated with a first user;

in response to receiving the notification, outputting an indication of the notification;

receiving one or more speech inputs;

in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

54. The method of item 53, wherein the notification of the event includes a representation of at least one of an incoming call, a reminder, a message, a voicemail, or a news alert.

55. The method of any of items 53 and 54, wherein receiving the notification of the event comprises:

receiving the notification from at least one of a second electronic device or a third electronic device.

56. The method of any of items 53-55, wherein outputting an indication of the notification comprises:

outputting the indication of the notification by the first electronic device or one of the additional electronic devices communicatively coupled to the first electronic device.

57. The method of any of items 53-56, wherein receiving the one or more speech inputs comprises:

receiving a first speech input regarding the notification;

outputting a response in accordance with the notification of the event; and

receiving a second speech input.

58. The method of any of items 53-57, wherein determining whether the notification is to be provided at the first electronic device comprises:

obtaining an identity of the user who provides at least one of the one or more speech inputs;

determining, based on the identity of the user who provides at least one of the one or more speech inputs and based on the notification, whether the notification is to be provided to the user who provides at least one of the one or more speech inputs; and

in accordance with a determination that the notification is to be provided to the user who provides at least one of the one or more speech inputs, determining whether the notification is to be provided at the first electronic device.

59. The method of item 58, wherein obtaining the identity of the user who provides at least one of the one or more speech inputs comprises:

obtaining authentication data associated with the user who provides at least one of the one or more speech inputs; and

obtaining a determination of the identity of the user who provides at least one of the one or more speech inputs based on the authentication data.

60. The method of item 59, wherein obtaining the determination of the identity of the user who provides at least one of the one or more speech inputs comprises:

providing the authentication data to at least one of a second electronic device and a third electronic device; and

receiving the identity of the user who provides at least one of the one or more speech inputs from at least one of the second electronic device and the third electronic device, wherein the identity of the user who provides at least one of the one or more speech inputs is determined based on the authentication data.

61. The method of any of items 53-60, wherein providing the notification at the first electronic device comprises:

providing an audio output associated with the notification at the first electronic device.

62. The method of any of items 53-61, further comprising:

in accordance with a determination that the notification is not to be provided at the first electronic device, determining an additional electronic device for providing the notification.

63. The method of item 62, wherein determining the additional electronic device for providing the notification is based on the one or more speech inputs.

64. The method of any of items 62 and 63, wherein determining the additional electronic device for providing the notification is based on context information.

65. The method of any of items 53-64, further comprising, prior to receiving the notification, establishing a connection between the first electronic device and at least one of a second electronic device or a third electronic device.

66. The method of item 65, wherein establishing the connection is based on a near-field communication between the first electronic device and the third electronic device.

67. The method of any of items 65 and 66, wherein establishing the connection is based on detecting of the third electronic device being within a predetermined distance from the first electronic device.

68. The method of any of items 65-67, wherein establishing the connection is based on a previous established connection between the first electronic device and the third electronic device.

69. The method of any of items 65-68, further comprising:

establishing a connection between the first electronic device and the third electronic device, wherein the third electronic device is associated with the first user; and

establishing a connection between the first electronic device and a fourth electronic device, wherein the fourth electronic device is associated with a second user.

70. The method of any of items 65-69, further comprising: after establishing a connection between the first electronic device and the third electronic device, notifying the second electronic device of the established connection.

71. The method of any of items 53-70, further comprising, establishing one or more connections between the first electronic device and one or more additional electronic devices, wherein the one or more additional electronic devices are the same type of devices as the first electronic device.

72. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors
of a first electronic device, cause the first electronic device to:
receive a notification of an event associated with a first user;
in response to receiving the notification, output an indication of the notification;
receive one or more speech inputs;
in accordance with the one or more speech inputs, determine whether the notification is to be provided at the first electronic device; and
in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

73. A first electronic device, comprising:
one or more processors;
memory; and
one or more programs stored in memory, the one or more programs including instructions for:
receiving a notification of an event associated with a first user;
in response to receiving the notification, outputting an indication of the notification;
receiving one or more speech inputs;

in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

74. An electronic device, comprising:

means for receiving a notification of an event associated with a first user;

in response to receiving the notification, means for outputting an indication of the notification;

means for receiving one or more speech inputs;

in accordance with the one or more speech inputs, means for determining whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, means for providing the notification at the first electronic device.

75. A first electronic device comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for performing the method of any one of items 53-71.

76. A first electronic device, comprising:

means for performing any of the methods of items 53-71.

77. A non-transitory computer-readable storage medium comprising one or more programs for execution by one or more processors of a first electronic device, the one or more programs including instructions which, when executed by the one or more processors, cause the first electronic device to perform the method of any one of items 53-71.

78. A system for operating a digital assistant, the system comprising means for performing any of the methods of items 53-71.

79. A method for providing a digital assistant service, comprising:

at a first electronic device with one or more processors and memory:

receiving, from a first user, a first speech input representing a user request;

obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

causing the second electronic device to provide at least a portion of the response to the user request.

80. The method of item 79, wherein obtaining the capability data comprises:

obtaining device capabilities associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device.

81. The method of item 80, wherein the device capabilities comprise one or more physical attributes associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device.

82. The method of any of items 80 and 81, wherein the device capabilities comprise data that are capable of being provided by the one of more electronic devices capable of being communicatively coupled to the first electronic device.

83. The method of any of items 79-82, wherein the first electronic device and the electronic devices capable of being communicatively coupled to the first electronic device are disposed within a predetermined boundary.

84. The method of any of items 79-83, wherein identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the first user comprises:

obtaining, based on the first speech input, one or more steps for responding to the user request; and

identifying, based on the capability data, the second electronic device for performing at least one step for responding to the user request.

85. The method of item 84, wherein obtaining one or more steps for responding to the user request comprises at least one of:

receiving, from a third electronic device remotely located from the first electronic device, a plan for responding to the user request; and

determining, by the first electronic device, a plan for responding to the user request, wherein the plan comprising one or more steps for responding to the user request.

86. The method of any of items 84 and 85, further comprising:

identifying, based on the capability data, one or more additional electronic devices for performing the remaining steps for responding the user request.

87. The method of any of items 79-86, wherein causing the second electronic device to provide at least a portion of the response to the user request comprises:

causing the second electronic device to perform at least one step for responding to the user request.

88. The method of item 87, further comprising:

causing one or more additional electronic devices to perform the remaining steps for responding to the user request.

89. The method of any of items 79-88, further comprising, prior to causing the second electronic device to provide at least a portion of the response to the user request:

providing a first speech output to the first user regarding providing at least a portion of the response by the second electronic device;

receiving a second speech input from the first user; and

in response to receiving the second speech input, determining whether at least a portion of the response is to be provided by the second electronic device.

90. The method of item 89, wherein providing a first speech output to the first user regarding providing at least a portion of the response at the second electronic device comprises:

annotating one or more steps for responding to the user request; and

providing the first speech output to the first user based on the annotation of one or more steps.

91. The method of any of items 79-90, further comprising, prior to obtaining capability data associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device:

establishing a connection between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

92. The method of item 91, wherein establishing the connection is based on a near-field communication between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

93. The method of any of items 91 and 92, wherein establishing the connection is based on detecting of the one or more electronic devices capable of being communicatively coupled to the first electronic device being within a predetermined distance from the first electronic device.

94. The method of any of items 91-93, wherein establishing the connection is based on one or more previous established connections between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

95. The method of any of items 79-94, further comprising, prior to obtaining capability data associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device:

inquiring the first user regarding accessing, by the first electronic device, the one or more electronic devices capable of being communicatively coupled to the first electronic device;

receiving a third speech input from the first user; and

in response to receiving the third speech input, determining whether the first electronic device is authorized to access the one or more electronic devices capable of being communicatively coupled to the first electronic device.

96. The method of item 95, further comprising:

providing one or more duration options for accessing the one or more electronic devices capable of being communicatively coupled to the first electronic device;

receiving a selection of a duration option from the first user; and

accessing the one or more electronic devices capable of being communicatively coupled to the first electronic device based on the selected duration option.

97. The method of any of items 79-96, further comprising, prior to causing the second electronic device to provide at least a portion of the response to the first user:

obtaining an identity of the first user ; and

determining, based on the identity of the first user, whether the first user is authorized to receive at least a portion of the response to the user request.

98. The method of any of items 97, wherein obtaining the identity of the first user comprises obtaining the identity based on a voice profile.

99. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors of a first electronic device, cause the first electronic device to:

receive, from a first user, a first speech input representing a user request;

obtain capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

in accordance with the capability data, identify, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

cause the second electronic device to provide at least a portion of the response to the first user.

100. A first electronic device, comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for:

receiving, from a first user, a first speech input representing a user request;

obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

causing the second electronic device to provide at least a portion of the response to the first user.

101. An electronic device, comprising:

means for receiving, from a first user, a first speech input representing a user request;

means for obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

means for in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

means for causing the second electronic device to provide at least a portion of the response to the first user.

102. A first electronic device comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for performing the method of any one of items 79-98.

103. A first electronic device, comprising:

means for performing any of the methods of items 79-98.

104. A non-transitory computer-readable storage medium comprising one or more programs for execution by one or more processors of a first electronic device, the one or more programs including instructions which, when executed by the one or more processors, cause the first electronic device to perform the method of any one of items 79-98.

105. A system for operating a digital assistant, the system comprising means for performing any of the methods of items 79-98.

[0363] The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many modifications and

variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the techniques and their practical applications. Others skilled in the art are thereby enabled to best utilize the techniques and various embodiments with various modifications as are suited to the particular use contemplated.

[0364] Although the disclosure and examples have been fully described with reference to the accompanying drawings, 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 claims.

[0365] As described above, one aspect of the present technology is the gathering and use of data available from various sources to obtain a user's identity. As described above, data for authenticating a user may include voice biometrics, facial recognition data, fingerprints, etc. The present disclosure contemplates that in some instances, this gathered data may include personal information data that uniquely identifies or can be used to contact or locate a specific person. As described above, informational capabilities of client devices may be obtained. The informational capabilities of client devices may include personal information data. Such personal information data can include personal identification data, demographic data, location-based data, telephone numbers, email addresses, home addresses, or any other identifying information.

[0366] The present disclosure recognizes that the use of such personal information data, in the present technology, can be used to the benefit of users. For example, the personal information data can be used to deliver targeted content that is of greater interest to the user. Accordingly, use of such personal information data enables calculated control of the delivered content. Further, other uses for personal information data that benefit the user are also contemplated by the present disclosure.

[0367] The present disclosure further contemplates that the entities responsible for the collection, analysis, disclosure, transfer, storage, or other use of such personal information data will comply with well-established privacy policies and/or privacy practices. In particular, such entities should implement and consistently use privacy policies and practices that are generally recognized as meeting or exceeding industry or governmental requirements for maintaining personal information data private and secure. For example, personal information from users

should be collected for legitimate and reasonable uses of the entity and not shared or sold outside of those legitimate uses. Further, such collection should occur only after receiving the informed consent of the users. Additionally, such entities would take any needed steps for safeguarding and securing access to such personal information data and ensuring that others with access to the personal information data adhere to their privacy policies and procedures. Further, such entities can subject themselves to evaluation by third parties to certify their adherence to widely accepted privacy policies and practices.

[0368] Despite the foregoing, the present disclosure also contemplates embodiments in which users selectively block the use of, or access to, personal information data. That is, the present disclosure contemplates that hardware and/or software elements can be provided to prevent or block access to such personal information data. For example, in the case of advertisement delivery services, the present technology can be configured to allow users to select to "opt in" or "opt out" of participation in the collection of personal information data during registration for services. In another example, users can select not to provide location information for targeted content delivery services. In yet another example, users can select to not provide precise location information, but permit the transfer of location zone information.

[0369] Therefore, although the present disclosure broadly covers use of personal information data to implement one or more various disclosed embodiments, the present disclosure also contemplates that the various embodiments can also be implemented without the need for accessing such personal information data. That is, the various embodiments of the present technology are not rendered inoperable due to the lack of all or a portion of such personal information data. For example, content can be selected and delivered to users by inferring preferences based on non-personal information data or a bare minimum amount of personal information, such as the content being requested by the device associated with a user, other non-personal information available to the content delivery services, or publically available information.

INTELLIGENT DIGITIAL ASSISTANT IN A MULTI-TASKING ENVIRONMENT

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Serial No. 62/348,728, entitled "INTELLIGENT DIGITAL ASSISTANT IN A MULTI-TASKING ENVIRONMENT," filed on June 10, 2016, which is hereby incorporated by reference in its entirety for all purposes.

FIELD

[0002] The present disclosure relates generally to a digital assistant and, more specifically, to a digital assistant that interacts with a user to perform a task in a multi-tasking environment.

BACKGROUND

[0003] Digital assistants are increasing popular. In a desktop or tablet environment, a user frequently multi-tasks including searching files or information, managing files or folders, playing movies or songs, editing documents, adjusting system configurations, sending emails, etc. It is often cumbersome and inconvenient for the user to manually perform multiple tasks in parallel and to frequently switch between tasks. It is thus desirable for a digital assistant to have the ability to assist the user to perform some of the tasks in a multi-tasking environment based on a user's voice input.

BRIEF SUMMARY

[0004] Some existing techniques for assisting the user to perform a task in a multi-tasking environment may include, for example, dictation. Typically, a user may be required to manually perform many other tasks in a multi-tasking environment. As an example, a user may have been working on a presentation yesterday on his or her desktop computer and may wish to continue to work on the presentation. The user is typically required to manually locate the presentation on his or her desktop computer, open the presentation, and continue the editing of the presentation.

[0005] As another example, a user may have been booking a flight on his or her smartphone when the user is away from his desktop computer. The user may wish to continue booking the flight when the desktop computer is available. In existing technologies, the user needs to launch a web browser and start over on the flight booking process at the user's desktop computer. In other words, the prior flight booking progress that the user made at the smartphone may not be continued at the user's desktop computer.

[0006] As another example, a user may be editing a document on his or her desktop computer and wish to change a system configuration such as changing the brightness level of the screen, turning on Bluetooth connections, or the like. In existing technologies, the user may need to stop editing the document, find and launch the brightness configuration application, and manually change the settings. In a multi-tasking environment, some existing technologies are incapable of performing tasks as described in the above examples based on a user's speech input. Providing a voice-enabled digital assistant in a multi-tasking environment is thus desired and advantageous.

[0007] Systems and processes for operating a digital assistant are provided. In accordance with one or more examples, a method includes, at a user device with one or more processors and memory, receiving a first speech input from a user. The method further includes identifying context information associated with the user device and determining a user intent based on the first speech input and the context information. The method further includes determining whether the user intent is to perform a task using a searching process or an object managing process. The searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device. The method further includes, in accordance with a determination that the user intent is to perform the task using the searching process, performing the task using the searching process. The method further includes, in accordance with the determination that the user intent is to perform the task using the object managing process, performing the task using the object managing process.

[0008] In accordance with one or more examples, a method includes, at a user device with one or more processors and memory, receiving a speech input from a user to perform a task. The

method further includes identifying context information associated with the user device and determining a user intent based on the speech input and context information associated with the user device. The method further includes, in accordance with user intent, determining whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device. The method further includes, in accordance with a determination that the task is to be performed at the user device and content for performing the task is located remotely, receiving the content for performing the task. The method further includes, in accordance with a determination that the task is to be performed at the first electronic device and the content for performing the task is located remotely to the first electronic device, providing the content for performing the task to the first electronic device.

[0009] In accordance with one or more examples, a method includes, at a user device with one or more processors and memory, receiving a speech input from a user to manage one or more system configurations of the user device. The user device is configured to concurrently provide a plurality of user interfaces. The method further includes identifying context information associated with the user device and determining a user intent based on the speech input and context information. The method further includes determining whether the user intent indicates an informational request or a request for performing a task. The method further includes, in accordance with a determination that the user intent indicates an informational request, providing a spoken response to the informational request. The method further includes, in accordance with a determination that the user intent indicates a request for performing a task, instantiating a process associated with the user device to perform the task.

[0010] Executable instructions for performing these functions are, optionally, included in a non-transitory computer-readable storage medium or other computer program product configured for execution by one or more processors. Executable instructions for performing these functions are, optionally, included in a transitory computer-readable storage medium or other computer program product configured for execution by one or more processors.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] For a better understanding of the various described embodiments, reference should be made to the Detailed Description below, in conjunction with the following drawings in which like reference numerals refer to corresponding parts throughout the figures.

[0012] FIG. 1 is a block diagram illustrating a system and environment for implementing a digital assistant according to various examples.

[0013] 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.

[0014] FIG. 2B is a block diagram illustrating exemplary components for event handling according to various examples.

[0015] FIG. 3 illustrates a portable multifunction device implementing the client-side portion of a digital assistant according to various examples.

[0016] FIG. 4 is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface according to various examples.

[0017] FIG. 5A illustrates an exemplary user interface for a menu of applications on a portable multifunction device according to various examples.

[0018] FIG. 5B illustrates an exemplary user interface for a multifunction device with a touch-sensitive surface that is separate from the display according to various examples.

[0019] FIG. 6A illustrates a personal electronic device according to various examples.

[0020] FIG. 6B is a block diagram illustrating a personal electronic device according to various examples.

[0021] FIG. 7A is a block diagram illustrating a digital assistant system or a server portion thereof according to various examples.

[0022] FIG. 7B illustrates the functions of the digital assistant shown in FIG. 7A according to various examples.

[0023] FIG. 7C illustrates a portion of an ontology according to various examples.

[0024] FIGs. 8A-8F illustrate functionalities of performing a task using a search process or an object managing process by a digital assistant according to various examples.

[0025] FIGs. 9A-9H illustrate functionalities of performing a task using a search process by a digital assistant according to various examples.

[0026] FIGs. 10A-10B illustrate functionalities of performing a task using an object managing process by a digital assistant according to various examples.

[0027] FIGs.11A-11D illustrate functionalities of performing a task using a search process by a digital assistant according to various examples.

[0028] FIGs. 12A-12D illustrate functionalities of performing a task using a search process or an object managing process by a digital assistant according to various examples.

[0029] FIGs. 13A-13C illustrate functionalities of performing a task using an object managing process by a digital assistant according to various examples.

[0030] FIGs. 14A-14D illustrate functionalities of performing a task at a user device using remotely located content by a digital assistant according to various examples.

[0031] FIGs. 15A-15D illustrate functionalities of performing a task at a first electronic device using remotely located content by a digital assistant according to various examples.

[0032] FIGs. 16A-16C illustrate functionalities of performing a task at a first electronic device using remotely located content by a digital assistant according to various examples.

[0033] FIGs. 17A-17E illustrate functionalities of performing a task at a user device using remotely located content by a digital assistant according to various examples.

[0034] FIGs. 18A-18F illustrate functionalities of providing system configuration information in response to an informational request of the user by a digital assistant according to various examples.

[0035] FIGs. 19A-19D illustrate functionalities of performing a task in response to a user request by a digital assistant according to various examples.

[0036] FIGs. 20A-20G illustrate a flow diagram of an exemplary process for operating a digital assistant according to various examples.

[0037] FIGs. 21A-21E illustrate a flow diagram of an exemplary process for operating a digital assistant according to various examples.

[0038] FIGs. 22A-22D illustrate a flow diagram of an exemplary process for operating a digital assistant according to various examples.

[0039] FIG. 23 illustrates a block diagram of an electronic device according to various examples.

Detailed Description

[0040] 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, of 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.

[0041] Techniques for providing a digital assistant in a multi-tasking environment are desirable. As described herein, techniques for providing a digital assistant in a multi-tasking environment are desired for various purposes such as reducing the cumbersomeness of searching objects or information, enabling efficient object management, maintaining continuity between tasks performed at the user device and at another electronic device, and reducing the user's manual effort in adjusting system configurations. Such techniques are advantageous by allowing the user to operate a digital assistant to perform various tasks using speech inputs in a multi-tasking environment. Further, such techniques alleviate the cumbersomeness or inconvenience

associated with performing various tasks in a multi-tasking environment. Furthermore, by allowing the user to perform tasks using speech, they are able to keep both hands on the keyboard or mouse while performing tasking that would require a context switch – effectively, allowing the digital assistant to perform tasks as if a "third-hand" of the user. As will be appreciated, by allowing the user to perform tasks using speech it allows the user to more efficiently complete tasks that may require multiple interactions with multiple applications. For example, searching for images and sending them to an individual in an email may require opening a search interface, entering search terms, selecting one or more results, opening am email for composition, copying or moving the resulting files to the open email, addressing the email and sending it. Such a task can be completed more efficiently by voice with a command such as "find pictures from X date and send them to my wife". Similar requests for moving files, searching for information on the internet, composing messages can all be made more efficient using voice, while simultaneously allowing the user to perform other tasks using their hands.

[0042] 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 storage could be termed a second storage, and, similarly, a second storage could be termed a first storage, without departing from the scope of the various described examples. The first storage and the second storage can both be storages and, in some cases, can be separate and different storages.

[0043] 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.

[0044] 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

[0045] 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.

[0046] 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.

[0047] 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.

[0048] 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.

[0049] 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 operate in a multi-tasking environment. A multi-tasking environment allows a user to operate device 104 to perform multiple tasks in parallel. For example, a multi-tasking environment may be a desktop or laptop environment, in which device 104 may perform one task in response to the user input received from a physical user-interface device and, in parallel, perform another task in response to the user's voice input. 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.

[0050] 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.

[0051] 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.

[0052] 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 devices 200, 400, or 600 described below with reference to FIGs. 2A, 4, and 6A-B. 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.

[0053] 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.

[0054] 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

[0055] 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.

[0056] 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).

[0057] 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.

[0058] 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.

[0059] 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.

[0060] 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 process 1200, 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 process 1200, 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.

[0061] 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.

[0062] 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.

[0063] 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).

[0064] 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).

[0065] 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.

[0066] 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.

[0067] Touch screen 212 has a touch-sensitive surface, sensor, or set of sensors that accept 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.

[0068] 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.

[0069] 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.

[0070] 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.

[0071] 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.

[0072] 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.

[0073] 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 or 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) and any other components associated with the generation, management, and distribution of power in portable devices.

[0074] 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.

[0075] 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.

[0076] 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).

[0077] 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.

[0078] 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.

[0079] 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.

[0080] 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.

[0081] 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.

[0082] 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.

[0083] 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).

[0084] 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.

[0085] 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.

[0086] 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.

[0087] 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.

[0088] 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, email 240, IM 241, browser 247, and any other application that needs text input).

[0089] 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).

[0090] 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) 264, 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.

[0091] 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.

[0092] 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.

[0093] 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.

[0094] 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.

[0095] 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.

[0096] 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;
- Email 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.

[0097] 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.

[0098] 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), email 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 email addresses to initiate and/or facilitate communications by telephone 238, video conference module 239, email 240, or IM 241; and so forth.

[0099] 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.

[0100] 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.

[0101] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, email client

module 240 includes executable instructions to create, send, receive, and manage email in response to user instructions. In conjunction with image management module 244, email client module 240 makes it very easy to create and send emails with still or video images taken with camera module 243.

[0102] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, and text input module 234, 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).

[0103] 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.

[0104] 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.

[0105] 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.

[0106] 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.

[0107] In conjunction with RF circuitry 208, touch screen 212, display controller 256, contact/motion module 230, graphics module 232, text input module 234, email 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.

[0108] 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).

[0109] 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).

[0110] 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.

[0111] 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.).

[0112] 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.

[0113] 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.

[0114] 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, email 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 email 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 email 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.

[0115] 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.

[0116] 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.

[0117] 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.

[0118] 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).

[0119] 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.

[0120] 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.

[0121] 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.

[0122] 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).

[0123] In some embodiments, event sorter 270 also includes a hit view determination module 272 and/or an active event recognizer determination module 273.

[0124] 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.

[0125] 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.

[0126] 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.

[0127] 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.

[0128] 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.

[0129] 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.

[0130] 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.

[0131] 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).

[0132] 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.

[0133] 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.

[0134] 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.

[0135] 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.

[0136] 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.

[0137] 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.

[0138] 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.

[0139] 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.

[0140] 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.

[0141] 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.

[0142] 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.

[0143] 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.

[0144] 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.

[0145] 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.

[0146] 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.

[0147] 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.

[0148] Attention is now directed towards embodiments of user interfaces that may be implemented on, for example, portable multifunction device 200.

[0149] 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:

[0150] 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 email client module 240, labeled "Mail," which optionally includes an indicator 510 of the number of unread emails;
    - 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.

[0151] 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.

[0152] 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.

[0153] 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.

[0154] Additionally, while the following examples are given primarily with reference to finger inputs (e.g., finger contacts, finger tap gestures, and/or 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.

[0155] 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.

[0156] 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.

[0157] 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.

[0158] 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.

[0159] 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.

[0160] 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 process 1200 (FIGs. 12A-D). 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. 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.

[0161] 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., link) may each constitute an affordance.

[0162] 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).

[0163] 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.

[0164] 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.

[0165] 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.

[0166] 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.

[0167] 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).

[0168] 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).

[0169] 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

[0170] 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.

[0171] 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.

[0172] 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).

[0173] 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 process 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).

[0174] 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) 712 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.

[0175] 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 process 1200, described below. One or more processors 704 can execute these programs, modules, and instructions, and read/write from/to the data structures.

[0176] 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.

[0177] 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 FIGs. 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.

[0178] 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.).

[0179] 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.

[0180] 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 731.

[0181] 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.

[0182] 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 a speech recognizer) for speech-to-text conversions.

[0183] 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.

[0184] 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.

[0185] 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'meɪɾoʊ/ and /tə'mɑtoʊ/. 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.

[0186] In some examples, the candidate pronunciations can be ranked based on the commonness of the candidate pronunciation. For example, the candidate pronunciation /tə'meɪɾoʊ/ can be ranked higher than /tə'mɑtoʊ/, 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'meɪɾoʊ/ can be associated with the United States, whereas the candidate pronunciation /tə'mɑtoʊ/ 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'meɪɾoʊ/ (associated with the United States) can be ranked higher than the candidate pronunciation /tə'mɑtoʊ/ (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).

[0187] 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'meɪroʊ/ 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."

[0188] In some examples, STT processing module 730 can use approximate matching techniques to determine words in a voice input. Thus, for example, the STT processing module 730 can determine that the sequence of phonemes /tə'meɪroʊ/ corresponds to the word "tomato," even if that particular sequence of phonemes is not one of the candidate sequence of phonemes for that word.

[0189] 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.

[0190] 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.

[0191] 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.

[0192] 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).

[0193] 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.

[0194] 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.

[0195] 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."

[0196] 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.

[0197] 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."

[0198] 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.

[0199] 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.

[0200] 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.

[0201] 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.

[0202] 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 speech input 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.

[0203] 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.

[0204] 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 speech inputs. 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 receive 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 receive 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.

[0205] 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.

[0206] 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.

[0207] 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.

[0208] 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.

[0209] 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.

[0210] 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.

[0211] 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.

[0212] 4. Exemplary Functions of a Digital Assistant – intelligent search and object management

[0213] FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C illustrate functionalities of performing a task using a search process or an object managing process by a digital assistant. In some examples, the digital assistant system (e.g., digital assistant system 700) is implemented by a user device according to various examples. In some examples, the user device, a server (e.g., server 108), or a combination thereof, may implement a digital assistant system (e.g., digital assistant system 700). The user device can be implemented using, for example, device 104, 200, or 400. In some examples, the user device is a laptop computer, a desktop computer, or a tablet computer. The user device can operate in a multi-tasking environment, such as a desktop environment.

[0214] With references to FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C, in some examples, a user device provides various user interfaces (e.g., user interfaces 810, 910, 1010, 1110, 1210, and 1310). The user device displays the various user interfaces on a

display (e.g., touch-sensitive display system 212, display 440) associated with the user device. The various user interfaces provide one or more affordances representing different processes (e.g., affordances 820, 920, 1020, 1120, 1220, and 1320 representing searching processes; and affordances 830, 930, 1030, 1130, 1230, and 1330 representing object managing processes). The one or more processes can be instantiated directly or indirectly by the user. For example, a user instantiates the one or more processes by selecting the affordances using an input device such as a keyboard, a mouse, a joystick, a finger, or the like. A user can also instantiate the one or more processes using a speech input, as described in more detail below. Instantiating a process includes invoking the process if the process is not already executing. If at least one instance of the process is executing, instantiating a process includes executing an existing instance of the process or generating a new instance of the process. For example, instantiating an object managing process includes invoking the object managing process, using an existing object managing process, or generate a new instance of the object managing process.

[0215] As shown in FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C, the user device displays, on a user interface (e.g., user interface 810, 910, 1010, 1110, 1210, and 1310) an affordance (e.g., affordance 840, 940, 1040, 1140, 1240, and 1340) to instantiate a digital assistant service. The affordance can be, for example, a microphone icon representing the digital assistant. The affordance can be displayed at any location on the user interfaces. For example, the affordance can be displayed on the dock (e.g., dock 808, 908, 1008, 1108, 1208, and 1308) at the bottom of the user interfaces, on the menu bar (e.g. menu bar 806, 906, 1006, 1106, 1206, and 1306) at the top of the user interfaces, in a notification center at the right side of the user interfaces, or the like. The affordance can also be displayed dynamically on the user interface. For example, the user device displays the affordance near an application user interface (e.g., an application window) such that the digital assistant service can be conveniently instantiated.

[0216] In some examples, the digital assistant is instantiated in response to receiving a pre-determined phrase. For example, the digital assistant is invoked in response to receiving a phrase such as “Hey, Assistant,” “Wake up, Assistant,” “Listen up, Assistant,” “OK, Assistant,” or the like. In some examples, the digital assistant is instantiated in response to receiving a selection of the affordance. For example, a user selects affordance 840, 940, 1040, 1140, 1240,

and/or 1340 using an input device such as a mouse, a stylus, a finger, or the like. Providing a digital assistant on a user device consumes computing resources (e.g., power, network bandwidth, memory, and processor cycles). In some examples, the digital assistant is suspended or shut down until a user invokes it. In some examples, the digital assistant is active for various periods of time. For example, the digital assistant can be active and monitoring the user's speech input during the time that various user interfaces are displayed, that the user device is turned on, that the user device is hibernating or sleeping, that the user is logged off, or a combination thereof.

[0217] With reference to FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C, a digital assistant receives one or more speech inputs, such as speech inputs 852, 854, 855, 856, 952, 954, 1052, 1054, 1152, 1252, or 1352, from a user. The user provides various speech inputs for the purpose of, for example, performing a task using a searching process or an object managing process. In some examples, the digital assistant receives speech inputs directly from the user at the user device or indirectly through another electronic device that is communicatively connected to the user device. The digital assistant receives speech inputs directly from the user via, for example, a microphone (e.g., microphone 213) of the user device. The user device includes a device that is configured to operate in a multi-tasking environment, such as a laptop computer, a desktop computer, a tablet, a server, or the like. The digital assistant can also receive speech inputs indirectly through one or more electronic devices such as a headset, a smartphone, a tablet, or the like. For instance, the user may speak to a headset (not shown). The headset receives the speech input from the user and transmits the speech input or a representation of it to the digital assistant of the user device via, for example, a Bluetooth connection between the headset and the user device.

[0218] With reference to FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C, in some embodiments, the digital assistant (e.g., represented by affordance 840, 940, 1040, 1140, 1240, and 1340) identifies context information associated with the user device. The context information includes, for example, user-specific data, metadata associated with one or more objects, sensor data, and user device configuration data. An object can be a target or a component of a process (e.g., an object managing process) associated with performing a task or a graphical element currently displayed on screen, and the object or graphical element may have or

may not currently have focus (e.g., be currently selected). For example, an object can include a file (e.g., a photo, a document), a folder, a communication (e.g., an email, a message, a notification, or a voicemail), a contact, a calendar, an application, an online resource, or the like. In some examples, the user-specific data includes log information, user preferences, the history of user's interaction with the user device, or the like. Log information indicates recent objects (e.g., a presentation file) used in a process. In some examples, metadata associated with one or more objects includes the title of the object, the time information of the object, the author of the object, the summary of the object, or the like. In some examples, the sensor data includes various data collected by a sensor associated with the user device. For example, the sensor data includes location data indicating the physical location of the user device. In some examples, the user device configuration data includes the current device configurations. For example, the device configurations indicate that the user device is communicatively connected to one or more electronic devices such as a smartphone, a tablet, or the like. As described in more detail below, the user device can perform one or more processes using the context information.

[0219] With reference to FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C, in response to receiving a speech input, the digital assistant determines a user intent based on the speech input. As described above, in some examples, the digital assistant processes a speech input via an I/O processing module (e.g., I/O processing module 728 as shown in FIG. 7B), an STT processing module (e.g., STT processing module 730 as shown in FIG. 7B), and a natural language processing module (e.g., natural language processing module 732 as shown in FIG. 7B). The I/O processing module forwards the speech input to an STT processing module (or a speech recognizer) for speech-to-text conversions. The speech-to-text conversion generates text based on the speech input. As described above, the STT processing module generates a sequence of words or tokens ("token sequence") and provides the token sequence to the natural language processing module. The natural language processing module performs natural language processing of the text and determines the user intent based on a result of the natural language processing. For example, the natural language processing module may attempt to associate the token sequence with one or more actionable intents recognized by the digital assistant. As described, once the natural language processing module identifies an actionable intent based on the user input, it generates a structured query to represent the identified actionable intent. The structured query includes one or more parameters associated with the

actionable intent. The one or more parameters are used to facilitate the performance of a task based on the actionable intent.

[0220] In some embodiments, the digital assistant further determines whether the user intent is to perform a task using a searching process or an object managing process. The searching process is configured to search data stored internally or externally to the user device. The object managing process is configured to manage objects associated with the user device. Various examples of determination of the user intent are provided below in more detail with respect to FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11D, 12A-12D, and 13A-13C.

[0221] With reference to FIG. 8A, in some examples, a user device receives a speech input 852 from a user to instantiate the digital assistant. Speech input 852 includes, for example, "Hey, Assistant." In response to the speech input, the user device instantiates the digital assistant represented by affordance 840 or 841 such that the digital assistant is actively monitoring subsequent speech inputs. In some examples, the digital assistant provides a spoken output 872 indicating that it is instantiated. For example, spoken output 872 includes "Go ahead, I am listening." In some examples, the user device receives a selection of affordance 840 or affordance 841 from the user to instantiate the digital assistant. The selection of affordance is performed by using an input device such as a mouse, a stylus, a finger, or the like.

[0222] With reference to FIG. 8B, in some examples, the digital assistant receives a speech input 854. Speech input 854 includes, for example, "Open the searching process and find the AAPL stock price today," or simply "show me the AAPL stock price today." Based on speech input 854, the digital assistant determines the user intent. For example, to determine the user intent, the digital assistant determines that the actionable intent is obtaining online information and that one or more parameters associated with this actionable intent include "AAPL stock price," and "today."

[0223] As described, in some examples, the digital assistant further determines whether the user intent is to perform a task using a searching process or an object managing process. In some embodiments, to make the determination, the digital assistant determines whether the speech input includes one or more keywords representing the searching process or the object managing process. For example, the digital assistant determines that speech input 854 includes keywords

or a phrase such as "open the searching process," indicating the user intent is to use the searching process to perform the task. As a result, the digital assistant determines that the user intent is to perform a task using the searching process.

[0224] As shown in FIG. 8B, in accordance with a determination the user intent is to perform the task using the searching process, the digital assistant performs the task using the searching process. As described, the natural language processing module of the digital assistant generates a structured query based on the user intent and passes the generated structured query to a task flow processing module (e.g., task flow processing module 736). The task flow processing module receives the structured query from the natural language processing module, completes the structured query, if necessary, and performs the actions required to "complete" the user's ultimate request. Performing the task using the searching process includes, for example, searching at least one object. In some embodiments, at least one object includes a folder, a file (e.g., a photo, an audio, a video), a communication (e.g., an email, a message, a notification, a voicemail), a contact, a calendar, an application (e.g., Keynote, Number, iTunes, Safari), an online informational source (e.g., Google, Yahoo, Bloomberg), or a combination thereof. In some examples, searching an object is based on metadata associated with the object. For example, the searching of a file or folder can use metadata such as a tag, a date, a time, an author, a title, a type of the file, a size, a page count, and/or a file location associated with the folder or file. In some examples, the file or folder is stored internally or externally to the user device. For example, the file or folder can be stored on the hard disk of the user device or stored on a cloud server. In some examples, searching a communication is based on metadata associated with the communication. For example, the searching of an email uses metadata such as the sender of the email, the receiver of the email, the sent/receive dates of the email, or the like.

[0225] As illustrated in FIG. 8B, in accordance with the determination that the user intent is to obtain the AAPL stock price using the searching process, the digital assistant performs the searching. For example, the digital assistant instantiates a searching process, represented by affordance 820, and causes the searching process to search today's AAPL stock price. In some examples, the digital assistant further causes the searching process to display a user interface 822

(e.g., a snippet or a window) providing text corresponding to speech input 854 (e.g., "Open the searching process and find the AAPL stock price today").

[0226] With reference to FIG. 8C, in some embodiments, the digital assistant provides a response based on a result of performing the task using the searching process. As illustrated in FIG. 8C, as a result of searching the AAPL stock price, the digital assistant displays a user interface 824 (e.g., a snippet or a window) providing the result of performing the task using the searching process. In some embodiments, user interface 824 is located within user interface 822 as a separate user interface. In some embodiments, user interfaces 824 and 822 are integrated together as a single user interface. On user interface 824, the search result of the stock price of AAPL is displayed. In some embodiments, user interface 824 further provides affordances 831 and 833. Affordance 831 enables closing of user interface 824. For example, if the digital assistant receives a user's selection of affordance 831, user interface 824 disappears or closes from the display of the user device. Affordance 833 enables moving or sharing of the search result displayed on user interface 824. For example, if the digital assistant receives the user's selection of affordance 833, it instantiates a process (e.g., the object managing process) to move or share user interface 824 (or the search result thereof) with a notification application. As shown in FIG. 8C, the digital assistant displays a user interface 826 that is associated with the notification application to provide the search result of AAPL stock price. In some embodiments, user interface 826 displays an affordance 827. Affordance 827 enables scrolling within user interface 826 such that the user can view the entire content (e.g., multiple notifications) within user interface 826 and/or indicates that relative position of the document with respect to its entire length and/or width. In some embodiments, user interface 826 displays results and/or dialog history (e.g., search results obtained from a current and/or past searching process) stored by the digital assistant. Further, in some examples, results of the performance the task are dynamically updated over time. For example, the AAPL stock price can be dynamically updated over time and displayed on user interface 826.

[0227] In some embodiments, the digital assistant also provides a spoken output corresponding to the search result. For example, the digital assistant (e.g., represented by affordance 840) provides a spoken output 874 including "Today's AAPL price is $100.00." In some examples, user interface 822 includes text corresponding to spoken output 874.

[0228] With reference to FIG. 8D, in some examples, the digital assistant instantiates a process (e.g., the object managing process) to move or share the search result displayed on user interface 824 in response to a subsequent speech input. For example, the digital assistant receives a speech input 855 such as "Copy the AAPL stock price to my notes." In response, the digital assistant instantiates a process to move or copy the search result (e.g., the AAPL stock price) to the user's note. As shown in FIG. 8D, in some examples, the digital assistant further displays a user interface 825 providing the copied or moved search result in user's note. In some examples, the digital assistant further provides a spoken output 875 such as "OK, the AAPL stock price is copied to your notes." In some examples, user interface 822 includes text corresponding to spoken output 875.

[0229] With reference to FIG. 8E, in some examples, the digital assistant determines that the user intent is to perform a task using the object managing process and performs the task using an object managing process. For example, the digital assistant receives a speech input 856 such as "Open the object managing process and show me all the photos from my Colorado trip," or simply "Show me all the photos from my Colorado trip." Based on speech input 856 and context information, the digital assistant determines the user intent. For example, the digital assistant determines that the actionable intent is to display photos and determines one or more parameters such as "all," and "Colorado trip." The digital assistant further determines which photos correspond to the user's Colorado trip using context information. As described, context information includes user-specific data, metadata of one or more objects, sensor data, and/or device configuration data. As an example, metadata associated with one or more files (e.g., file 1, file 2, and file 3 displayed in user interface 832) indicates that the file names includes the word "Colorado" or a city name of Colorado (e.g., "Denver"). The metadata may also indicate that a folder name includes the word "Colorado" or a city name of Colorado (e.g., "Denver"). As another example, sensor data (e.g., GPS data) indicates that the user was travelling within Colorado during a certain period of time. As a result, any photos the user took during that particular period of time are photos taken during the user's Colorado trip. As well, photos themselves may include geotagged metadata that associates the photo with the location at which it was taken. Based on the context information, the digital assistant determines that the user intent is to, for example, display photos stored in a folder having a folder name "Colorado trip," or display photos taken during the period of time that the user was travelling within Colorado.

[0230] As described, in some examples, the digital assistant determines whether the user intent is to perform a task using a searching process or an object managing process. To make such determination, the digital assistant determines whether the speech input includes one or more keywords representing the searching process or the object managing process. For example, the digital assistant determines that speech input 856 includes keywords or a phrase such as "open the object managing process," indicating that the user intent is to use the object managing process to perform the task.

[0231] In accordance with a determination the user intent is to perform the task using the object managing process, the digital assistant performs the task using the object managing process. For example, the digital assistant searches at least one object using the object managing process. In some examples, at least one object includes at least one of a folder or a file. A file can include at least one of a photo, an audio (e.g., a song), or a video (e.g., a movie). In some examples, searching a file or a folder is based on metadata associated with the folder or file. For example, the searching of a file or folder uses metadata such as a tag, a date, a time, an author, a title, a type of the file, a size, a page count, and/or a file location associated with the folder or file. In some examples, the file or folder can be stored internally or externally to the user device. For example, the file or folder can be stored on the hard disk of the user device or stored on a cloud server.

[0232] As illustrated in FIG. 8E, in accordance with the determination that the user intent is, for example, to display photos stored in a folder having a folder name "Colorado trip," or display photos taken during the period of time that the user was travelling within Colorado, the digital assistant performs the task using the object managing process. For example, the digital assistant instantiates an object managing process represented by affordance 830 and causes the object managing process to search for photos from the user's Colorado trip. In some examples, the digital assistant also causes the object managing process to display a snippet or a window (not shown) providing text of the user's speech input 856.

[0233] With reference to FIG. 8F, in some embodiments, the digital assistant further provides a response based on a result of performing the task using the object managing process. As illustrated in FIG. 8F, as a result of searching the photos of the user's Colorado trip, the

digital assistant displays a user interface 834 (e.g., a snippet or a window) providing the result of performing the task using the object managing process. For example, on user interface 834, a preview of the photos is displayed. In some examples, the digital assistant instantiates a process (e.g., the object managing process) to perform additional tasks on the photos, such as inserting the photos to a document or attaching the photos to email. As described in more detail below, the digital assistant can instantiate a process to perform the additional tasks in response to a user's additional speech input. As well, the digital assistant can perform multiple tasks in response to a single speech input, such as "send the photos from my Colorado trip to my Mom by email." The digital assistant can also instantiate a process to perform such additional tasks in response to the user's input using an input device (e.g., a mouse input to select of one or more affordances or perform a drag-and-drop operation). In some embodiments, the digital assistant further provides a spoken output corresponding to the result. For example, the digital assistant provides a spoken output 876 including "Here are the photos from your Colorado trip."

[0234] With reference to FIG. 9A, in some examples, user's speech input may not include one or more keywords indicating whether the user intent is to use the searching process or the object managing process. For example, the user provides a speech input 952 such as "What is the score of today's Warriors game?" Speech input 952 does not include keywords indicating "the searching process" or the "object managing process." As a result, the keywords may not be available for the digital assistant to determine whether the user intent is to perform the task using the searching process or the object managing process.

[0235] In some embodiments, to determine whether the user intent is to perform the task using the searching process or the object managing process, the digital assistant determines whether the task is associated with searching based on the speech input. In some examples, a task that is associated with searching can be performed by either the searching process or the object managing process. For example, both the searching process and the object managing process can search a folder and a file. In some examples, the searching process can further search a variety of objects including online information sources (e.g., websites), communications (e.g., emails), contacts, calendars, or the like. In some examples, the object managing process may not be configured to search certain objects such as online information sources.

[0236] In accordance with a determination that the task is associated with searching, the digital assistant further determines whether performing the task requires the searching process. As described, if a task is associated with searching, either the searching process or the object managing process can be used to perform the task. However, the object managing process may not be configured to search certain objects. As a result, to determine whether the user intent is to use the searching process or the object managing process, the digital assistant further determines whether the task requires the searching process. For example, as illustrated in FIG. 9A, based on speech input 952, the digital assistant determines that the user intent is, for example, to obtain the score of today's Warriors game. According to the user intent, the digital assistant further determines that performing the task requires searching online information sources and therefore is associated with searching. The digital assistant further determines whether performing the task requires the searching process. As described, in some examples, the searching process is configured to search online information sources such as websites, while the object managing process may not be configured to search such online information sources. As a result, the digital assistant determines that searching online information sources (e.g., searching Warriors' website to obtain the score) requires the searching process.

[0237] With reference to FIG. 9B, in some embodiments, in accordance with a determination that performing the task requires the searching process, the digital assistant performs the task using the searching process. For example, in accordance with the determination that searching the score of today's Warriors game requires the searching process, the digital assistant instantiates a searching process represented by affordance 920, and causes the searching process to search score of today's Warriors game. In some examples, the digital assistant further causes the searching process to display a user interface 922 (e.g., a snippet or a window) providing text of user speech input 952 (e.g., "What is the score of today's Warriors game?"). User interface 922 includes one or more affordances 921 and 927. Similar to described above, affordance 921(e.g., a close button) enables closing of user interface 922 and affordance 927 (e.g., a scrolling bar) enables scrolling within user interface 922 such that the user can view the entire content within user interface 922.

[0238] With reference to FIG. 9B, in some examples, based on the search results, the digital assistant further provides one or more responses. As illustrated in FIG. 9B, as a result of

searching the score of today's Warriors game, the digital assistant displays a user interface 924 (e.g., a snippet or a window) providing the result of performing the task using the searching process. In some embodiments, user interface 924 is located within user interface 922 as a separate user interface. In some embodiments, user interfaces 924 and 922 are integrated together as a single user interface. In some examples, the digital assistant displays the user interface 924 providing the current search results (e.g., the Warriors game score) together with another user interface (e.g., user interface 824 shown on FIG. 8C) providing prior search results (e.g., the AAPL stock price). In some embodiments, the digital assistant only displays user interface 924 providing the current search results and does not display another user interface providing prior search results. As illustrated in FIG. 9B, the digital assistant only displays user interface 924 to provide the current search results (e.g., the Warriors game score). In some examples, affordance 927 (e.g., a scrolling bar) enables scrolling within user interface 922 such that the user can view the prior search results. Further, in some examples, prior search results dynamically update or refresh, e.g., such that stock prices, sports score, weather forecast, etc., update over time.

[0239] As illustrated in FIG. 9B, on user interface 924, the search result of the score of today's Warriors game is displayed (e.g., Warriors 104-89 Cavaliers). In some embodiments, user interface 924 further provides affordances 923 and 925. Affordance 923 enables closing of user interface 924. For example, if the digital assistant receives a user's selection of affordance 923, user interface 924 disappears or closes from the display of the user device. Affordance 925 enables moving or sharing of the search result displayed on user interface 924. For example, if the digital assistant receives the user's selection of affordance 925, it moves or shares user interface 924 (or the search result thereof) with a notification application. As shown in FIG. 9B, the digital assistant displays user interface 926 that is associated with the notification application to provide the search result of Warriors game score. As described, results of the performance the task are dynamically updated over time. For example, the Warriors game score can be dynamically updated over time while the game is ongoing and displayed on user interface 924 (e.g., the snippet or window) and/or on user interface 926 (e.g., the notification application user interface). In some embodiments, the digital assistant further provides a spoken output corresponding to the search result. For example, the digital assistant represented by affordance 940 or 941 provides a spoken output 972 such as "Warriors beats Cavaliers, 104-89." In some

examples, user interface 922 (e.g., a snippet or a window) provides text corresponding to spoken output 972.

[0240] As described above, in some embodiments, the digital assistant determines whether the task is associated with searching, and in accordance with such a determination, the digital assistant determines whether performing the task requires the searching process. With reference to FIG. 9C, in some embodiments, the digital assistant determines that performing the task does not require the searching process. For example, as illustrated in FIG. 9C, the digital assistant receives a speech input 954 such as "Show me all the files called Expenses." Based on speech input 954 and context information, the digital assistant determines that user intent is to display all the files having the word "Expenses" (or a portion, a variation, a paraphrase thereof) contained in their file names, the metadata, the content of the files, or the like. According to the user intent, the digital assistant determines that the task to be performed includes searching all the files associated with the word "Expenses." As a result, the digital assistant determines that performing the task is associated with searching. As described above, in some examples, the searching process and the object managing process can both perform searching of files. As a result, the digital assistant determines that performing the task of searching all the files associated with the word "Expenses" does not require the searching process.

[0241] With reference to FIG. 9D, in some examples, in accordance with a determination that performing the task does not require the searching process, the digital assistant determines, based on a pre-determined configuration, whether the task is to be performed using the searching process or the object managing process. For example, if both the searching process and the object managing process can perform the task, a pre-determined configuration may indicate that the task is to be performed using the searching process. The pre-determined configuration can be generated and updated using context information such as user preferences or user-specific data. For example, the digital assistant determines that historically, for a particular user, the searching process was selected more frequently than the object managing process for file searching. As a result, the digital assistant generates or updates the pre-determined configuration to indicate that the searching process is the default process for searching files. In some examples, the digital assistant generates or updates the pre-determined configuration to indicate that the object managing process is the default process.

[0242] As illustrated in FIG. 9D, based on a pre-determined configuration, the digital assistant determines that the task of searching all the files associated with the word "Expense" is to be performed using the searching process. As a result, the digital assistant performs the searching of all the files associated with the word "Expenses" using the searching process. For example, the digital assistant instantiates a searching process represented by affordance 920 displayed on user interface 910, and causes the searching process to search all files associated with the word "Expenses." In some examples, the digital assistant further provides a spoken output 974, informing the user that the task is being performed. Spoken output 974 includes, for example, "OK, searching all files called 'Expenses'." In some examples, the digital assistant further causes the searching process to display a user interface 928 (e.g., a snippet or a window) providing text corresponding to speech input 954 and spoken output 974.

[0243] With reference to FIG. 9E, in some embodiments, the digital assistant further provides one or more responses based on a result of performing the task using the searching process. As illustrated in FIG. 9E, as a result of searching all files associated with the word "Expenses," the digital assistant displays a user interface 947 (e.g., a snippet or a window) providing the search results. In some embodiments, user interface 947 is located within user interface 928 as a separate user interface. In some embodiments, user interfaces 947 and 928 are integrated together as a single user interface. On user interface 947, a list of files that are associated with the word "Expenses" are displayed. In some embodiments, the digital assistant further provides a spoken output corresponding to the search result. For example, the digital assistant represented by affordance 940 or 941 provides a spoken output 976 such as "Here are all the files called Expenses." In some examples, the digital assistant further provides, on user interface 928, text corresponding to spoken output 976.

[0244] In some embodiments, the digital assistant provides one or more links associated with the result of performing the task using the searching process. A link enables instantiating a process (e.g., opening a file, invoking an object managing process) using the search result. As illustrated in FIG. 9E, on user interface 947, the list of files (e.g., Expenses File 1, Expenses File 2, Expenses File 3) represented by their file names can be associated with links. As an example, a link is displayed on the side of each file name. As another example, the file names is displayed in a particular color (e.g., blue) indicating that the file names are associated with links. In some

examples, the file names associated with links are displayed in the same color as other items displayed on user interface 947.

[0245] As described, a link enables instantiating a process using the search result. Instantiating a process includes invoking the process if the process is not already running. If at least one instance of the process is running, instantiating a process includes executing an existing instance of the process or generating a new instance of the process. For example, instantiating an object managing process includes invoking the object managing process, using an existing object managing process, or generating a new instance of the object managing process. As illustrated in FIGs. 9E and 9F, a link displayed on user interface 947 enables managing an object (e.g., a file) associated with the link. For example, user interface 947 receives a user selection of a link (e.g., a selection by a cursor 934) associated with a file (e.g., "Expenses file 3"). In response, the digital assistant instantiates an object managing process represented by affordance 930 to enable managing of the file. As shown in FIG. 9F, the digital assistant displays a user interface 936 (e.g., a snippet or a window) providing the folder containing the file associated with the link (e.g., "Expenses file 3"). Using user interface 936, the digital assistant instantiates the object managing process to perform one or more additional tasks (e.g., copying, editing, viewing, moving, compressing, or the like) with respect to the files.

[0246] With reference back to FIG. 9E, in some examples, a link displayed on user interface 947 enables direct viewing and/or editing of the object. For example, the digital assistant, via user interface 947, receives a selection of a link (e.g., a selection by a cursor 934) associated with a file (e.g., "Expenses file 3"). In response, the digital assistant instantiates a process (e.g., a document viewing/editing process) to view and/or edit the file. In some examples, the digital assistant instantiates the process to view and/or edit the file without instantiating an object managing process. For example, the digital assistant directly instantiates a Number process or an Excel process to view and/or edit of the Expense file 3.

[0247] With reference to FIGs. 9E and 9G, in some examples, the digital assistant instantiates a process (e.g., the searching process) to refine the search results. As illustrated in FIGs. 9E and 9G, the user may desire to refine the search result displayed on user interface 947. For example, the user may desire to select one or more files from the search results. In some

examples, the digital assistant receives, from the user, a speech input 977 such as "Just the ones Kevin sent me that I tagged with draft." Based on speech input 977 and context information, the digital assistant determines that the user intent is to display only the Expenses files that were sent from Kevin and that are associated with draft tags. Based on the user intent, the digital assistant instantiates a process (e.g., the searching process) to refine the search results. For example, as shown in FIG. 9G, based on the search result, the digital assistant determines that Expenses File 1 and Expense file 2 were sent from Kevin to the user and were tagged. As a result, the digital assistant continues to display these two files on user interface 947 and remove the Expense file 3 from user interface 947. In some examples, the digital assistant provides a spoken output 978 such as "Here are just the ones Kevin sent you that you tagged with draft." The digital assistant may further provide text corresponding to spoken output 978 on user interface 928.

[0248] With reference to FIG. 9H, in some examples, the digital assistant instantiates a process (e.g., an object managing process) to perform an object managing task (e.g., coping, moving, sharing, etc.). For example, as shown in FIG. 9H, the digital assistant receives, from the user, a speech input 984 such as "Move the Expenses file 1 to Documents folder." Based on speech input 984 and context information, the digital assistant determines that the user intent is to copy or move Expense file 1 from its current folder to Document folder. In accordance with the user intent, the digital assistant instantiates a process (e.g., the object managing process) to copy or move Expense file 1 from its current folder to Document folder. In some examples, the digital assistant provides a spoken output 982 such as "Ok, moving Expenses File 1 to your Documents folder." In some examples, the digital assistant furthers provide text corresponding to spoken output 982 on user interface 928.

[0249] As described, in some examples, a user's speech input may not include keywords indicating whether the user intent is to perform the task using the search process or the object managing process. With reference to FIG. 10A-10B, in some embodiments, the digital assistant determines that performing the task does not require the searching process. In accordance with the determination, the digital assistant provides a spoken output requesting the user to select the searching process or the object managing process. For example, as shown in FIG. 10A, the digital assistant receives, from the user, a speech input 1052 such as "Show me all the files called 'Expenses.'" Based on speech input 1052 and context information, the digital assistant

determines that the user intent is to display all the files associated with the word "Expense." In accordance with the user intent, the digital assistant further determines that the task can be performed by either the searching process or the object managing process, and therefore does not require the search process. In some examples, the digital assistant provides a spoken output 1072 such as "Do you want to search using the searching process or the object managing process?" In some examples, the digital assistant receives, from the user, a speech input 1054 such as "Object managing process." Speech input 1054 thus indicates that the user intent is to perform the task using the object managing process. According to the selection, for example, the digital assistant instantiates an object managing process represented by affordance 1030 to search all the files associated with the word "Expenses." As shown in FIG. 10B, similar to those described above, as a result of the searching, the digital assistant displays a user interface 1032 (e.g., a snippet or a window) providing a folder containing the files associated with the word "Expenses". Similar to those described above, using user interface 1032, the digital assistant instantiates the object managing process to perform additional one or more tasks (e.g., copying, editing, viewing, moving, compressing, or the like) with respect to the files.

[0250] With reference to FIGs. 11A and 11B, in some embodiments, the digital assistant identifies context information and determines the user intent based on the context information and the user's speech input. As illustrated in FIG. 11A, the digital assistant represented by affordance 1140 or 1141 receives a speech input 1152 such as "Open the Keynote presentation I created last night." In response to receiving speech input 1152, the digital assistant identifies context information such as the history of the user's interaction with the user device, the metadata associated with files that the user recently worked on, or the like. For example, the digital assistant identifies the metadata such as the date, the time, and the type of files the user worked on yesterday from 6p.m.-2a.m. Based on the identified context information and speech input 1152, the digital assistant determines that the user intent includes searching a Keynote presentation file associated with metadata indicating that the file was edited approximately 6p.m.-12a.m yesterday; and instantiating a process (e.g., a Keynote process) to open the presentation file.

[0251] In some examples, the context information includes application names or identifications (IDs). For example, a user's speech input provides "Open the Keynote

presentation," "find my Pages document," or "find my HotNewApp documents." The context information includes the application names (e.g., Keynote, Pages, HotNewApp) or application IDs. In some examples, the context information is dynamically updated or synchronized. For example, the context information is updated in real time after the user installs a new application named HotNewApp. In some examples, the digital assistant identifies the dynamically updated context information and determines the user intent. For example, the digital assistant identifies the application names Keynote, Pages, HotNewApp or their IDs and determines the user intent according to the application names/IDs and speech inputs.

[0252] In accordance with the user intent, the digital assistant further determines whether the user intent is to perform the task using the searching process or the object managing process. As described, the digital assistant makes such determination based on one or more keywords included in the speech input, based on whether the task requires the searching process, based on a pre-determined configuration, and/or based on the user's selection. As illustrated in FIG. 11A, speech input 1152 does not include keywords that indicate whether the user intent is to use the searching process or the object managing process. As a result, the digital assistant determines, for example, based on a pre-determined configuration that the user intent is to use the object managing process. In accordance with the determination, the digital assistant instantiate an object managing process to search a Keynote presentation file associated with metadata that indicates the file was edited approximately 6p.m.-12a.m yesterday. In some embodiments, the digital assistant further provides a spoken output 1172 such as "OK, looking for the Keynote presentation you created last night."

[0253] In some embodiments, context information is used in performing the task. For example, application names and/or IDs can be used to form a query for searching the application and/or objects (e.g., files) associated with the application names/IDs. In some examples, a server (e.g., server 108) forms a query using the application names (e.g., Keynote, Pages, HotNewApp) and/or IDs and sends the query to the digital assistant of a user device. Based on the query, the digital assistant instantiates a searching process or an object managing process to search one or more applications and/or objects. In some examples, the digital assistant only searches the objects (e.g., files) that correspond to the application name/ID. For example, if a query includes an application name "Pages," the digital assistant only searches Pages files and does not search

other files (e.g., Word files) that can be opened by a Pages application. In some examples, the digital assistant searches all objects that is associated with the application name/ID in the query.

[0254] With references to FIGs. 11B and 11C, in some embodiments, the digital assistant provides one or more responses in accordance with a confidence level associated with the results of performing the task. Inaccuracies may exist or arise during the determination of the user intent, the determination of whether the user intent is to perform the task using the searching process or the object managing process, and/or the performance of the task. In some examples, the digital assistant determines a confidence level representing the accuracy of determining the user intent based on the speech input and context information, the accuracy of determining whether the user intent is to perform the task using the searching process or the object managing process, the accuracy of performing the task using the searching process or the object managing process, or a combination thereof.

[0255] Continuing the above example illustrated in FIG. 11A, based on speech input 1152 such as "Open the Keynote presentation I created last night," the digital assistant instantiates an object managing process to perform a search of a Keynote presentation file associated with metadata that indicates the file was edited approximately 6p.m.-12a.m yesterday. The search result may include a single file that fully matches the search criteria. That is, the single file is a presentation file that was edited approximately 6p.m.-12a.m yesterday. Accordingly, the digital assistant determines that the accuracy of the search is high and thus determines that the confidence level is high. As another example, the search result may include a plurality of files that partially match the search criteria. For instance, no file is a presentation file that was edited approximately 6p.m.-12a.m yesterday, or multiple files are presentation files that were edited approximately 6p.m.-12a.m yesterday. Accordingly, the digital assistant determines that the accuracy of the search is medium or low and thus determines that the confidence level is medium or low.

[0256] As illustrated in FIGs. 11B and 11C, the digital assistant provides a response in accordance with the determination of the confidence level. In some examples, the digital assistant determines whether the confidence level is greater than or equal to a threshold confidence level. In accordance with a determination that the confidence level is greater than or

equal to the threshold confidence level, the digital assistant provides a first response. In accordance with a determination that the confidence level is less than a threshold confidence level, the digital assistant provides a second response. In some examples, the second response is different from the first response. As shown in FIG. 11B, if the digital assistant determines that the confidence level is greater than or equal to a threshold confidence level, the digital assistant instantiates a process (e.g., a Keynote process represented by user interface 1142) to enable the viewing and editing of the file. In some examples, the digital assistant provides a spoken output such as "Here is the presentation you created last night," and displays the text of the spoken output in a user interface 1143. As shown in FIG. 11C, if the digital assistant determines that the confidence level is less than a threshold confidence level, the digital assistant displays a user interface 1122 (e.g., a snippet or a window) providing a list of candidate files. Each of the candidate files may partially satisfy the search criteria. In some embodiments, the confidence level can be pre-determined and/or dynamically updated based on user preferences, historical accuracy rates, or the like. In some examples, the digital assistant further provides a spoken output 1174 such as "Here are all the presentations created last night," and displays the text corresponding to spoken output 1174 on user interface 1122.

[0257] With reference to FIG. 11D, in some embodiments, the digital assistant instantiates a process (e.g., the Keynote presentation process) to perform additional tasks. Continuing with the above example, as shown in FIGs. 11B and 11D, the user may desire to display the presentation file in a full screen mode. The digital assistant receives, from the user, a speech input 1154 such as "Make it full screen." Based on speech input 1154 and context information, the digital assistant determines that the user intent is to display the presentation file in a full screen mode. In accordance with the user intent, the digital assistant causes the Keynote presentation process to display the slides in a full-screen mode. In some examples, the digital assistant provides a spoken output 1176 such as "OK, showing your presentation in full screen."

[0258] With reference to FIGs. 12A-12C, in some embodiments, the digital assistant determines, based on a single speech input or an utterance, that the user intent is to perform a plurality of tasks. In accordance with the user intent, the digital assistant further instantiates one or more processes to perform the plurality of tasks. For example, as shown in FIG. 12A, the digital assistant represented by affordance 1240 or 1241 receives a single speech input 1252 such

as "Show me all the photos from my Colorado trip, and send them to my mom." Based on speech input 1252 and context information, the digital assistant determines that the user intent is to perform a first task and a second task. Similar to those described above, the first task is to display photos stored in a folder having a folder name "Colorado trip," or display photos taken during the period of time that the user is travelling within Colorado. With respect to the second task, the context information may indicate that a particular email address stored in the user's contacts is tagged as the user's mom. Accordingly, the second task is to send an email containing the photos associated with the Colorado trip to the particular email address.

[0259] In some examples, the digital assistant determines, with respect to each task, whether the user intent is to perform the task using the searching process or the object managing process. As an example, the digital assistant determines that the first task is associated with searching and the user intent is to perform the first task using the object managing process. As illustrated in FIG. 12B, in accordance with a determination the user intent is to perform the first task using the object managing process, the digital assistant instantiates the object managing process to search photos associated with the user's Colorado trip. In some examples, the digital assistant displays a user interface 1232 (e.g., a snippet or a window) providing a folder including the search result (e.g., photos 1, 2, and 3). As another example, the digital assistant determines that the first task is associated with searching and the user intent is to perform the first task using the searching process. As illustrated in FIG. 12C, in accordance with a determination the user intent is to perform the first task using the searching process, the digital assistant instantiates the searching process to search photos associated with the user's Colorado trip. In some examples, the digital assistant displays a user interface 1234 (e.g., a snippet or a window) providing photos and/or links associated with the search result (e.g., photos 1, 2, and 3).

[0260] As another example, the digital assistant determines that the second task (e.g., sending an email containing the photos associated with the Colorado trip to the particular email address) is not associated with searching or associated with managing an object. In accordance with the determination, the digital assistant determines whether the task can be performed using a process that is available to the user device. For example, the digital assistant determines that the second task can be performed using an email process at the user device. In accordance with the determination, the digital assistant instantiates the process to perform the second task. As

illustrated in FIGs. 12B and 12C, the digital assistant instantiates the email process and displays user interfaces 1242 and 1244 associated with the email process. The email process attaches the photos associated with the user's Colorado trip to email messages. As shown in FIGs. 12B and 12C, in some embodiments, the digital assistant further provides spoken outputs 1272 and 1274 such as "Here are the photos from your Colorado trip. I am ready to send the photos to your mom, proceed?" In some examples, the digital assistant displays text corresponding to spoken output 1274 on user interface 1244. In response to spoken outputs 1272 and 1274, the user provides a speech input such as "OK." Upon receiving the speech input from the user, the digital assistant causes the email process to send out the email messages.

[0261] Techniques for performing a plurality of tasks based on multiple commands contained within a single speech input or an utterance may be found, for example, in related applications: U.S. Patent Application Serial No. 14/724,623, titled "MULTI-COMMAND SINGLE UTTERANCE INPUT METHOD," filed May 28, 2015, which claims the benefit of priority of U.S. Provisional Patent Application No. 62/005,556, entitled "MULTI-COMMAND SINGLE UTTERANCE INPUT METHOD," filed on May 30, 2014; and U.S. Provisional Patent Application No. 62/129,851, entitled "MULTI-COMMAND SINGLE UTTERANCE INPUT METHOD," filed on March 8, 2015. Each of these applications is hereby incorporated by reference in their entirety.

[0262] As illustrated in FIGs. 12C and 12D, in some examples, the digital assistant causes a process to perform additional tasks based on the user's additional speech inputs. For example, in view of the search result displayed in user interface 1234, the user may desire to send some, but not all, of the photos. The user provides a speech input 1254 such as "Send only Photo 1 and Photo 2." In some examples, the digital assistant receives speech input 1254 after the user selects affordance 1235 (e.g., a microphone icon displayed on user interface 1234). The digital assistant determines, based on speech input 1254 and context information, that the user intent is to send an email attaching only Photo 1 and Photo 2. In accordance with the user intent, the digital assistant causes the email process to remove Photo 3 from the email message. In some examples, the digital assistant provides a spoken output 1276, such as "OK, attaching Photo 1 and Photo 2 to your email," and displays the text corresponding to spoken output 1276 on user interface 1234.

[0263] With reference to FIG. 13A, in some embodiments, in accordance with a determination that the task is not associated with searching, the digital assistant determines whether the task is associated with managing at least one object. As illustrated in FIG. 13A, for example, the digital assistant receives a speech input 1352 such as "Create a new folder on the desktop called Projects." Based on speech input 1352 and context information, the digital assistant determines that the user intent is to generate a new folder at the desktop with a folder name "Projects." The digital assistant further determines that the user intent is not associated with searching, and instead is associated with managing an object (e.g., a folder). Accordingly, the digital assistant determines that the user intent is to perform a task using the object managing process.

[0264] In some examples, in accordance with the determination that the user intent is to perform the task using the object managing process, the digital assistant performs the task using the object managing process. Performing the task using the object managing process can include, for example, creating at least one object (e.g., creating a folder or a file), storing at least one object (e.g., storing a folder, a file, or a communication), and compressing at least one object (e.g., compressing folders and files). Performing the task using the object managing process can further include, for example, copying or moving at least one object from a first physical or virtual storage to a second physical or virtual storage. For instance, the digital assistant instantiates an object managing process to cut and paste a file from the user device to a flash drive or a cloud drive.

[0265] Performing the task using the object managing process can further include, for example, deleting at least one object stored in a physical or virtual storage (e.g., deleting a folder or a file) and/or recovering at least one object stored at a physical or virtual storage (e.g., recovering a deleted folder or a deleted file). Performing the task using the object managing process can further include, for example, marking at least one object. In some examples, marking of an object can be visible or invisible. For example, the digital assistant can cause the object managing process to generate a "like" sign for a social media post, to tag an email, to mark a file, or the like. The marking may be visible by displaying, for example, a flag, a sign, or the like. The marking may also be performed with respect to the metadata of the object such that

a storage (e.g., a memory) content of the metadata is varied. The metadata may or may not be visible.

[0266] Performing the task using the object managing process can further include, for example, backing up at least one object according to a predetermined time period for backing up or upon the user's request. For example, the digital assistant can cause the object managing process to instantiate a backup program (e.g., time machine program) to backup folders and files. The backup can be performed automatically according to a pre-determined schedule (e.g., once a day, a week, a month, or the like) or according to a user request.

[0267] Performing the task using the object managing process can further include, for example, sharing at least one object among one or more electronic devices communicatively connected to the user device. For example, the digital assistant can cause the object managing process to share a photo stored on the user device with another electronic device (e.g., the user's smartphone or tablet).

[0268] As illustrated in FIG. 13B, in accordance with the determination that the user intent is to perform the task using the object managing process, the digital assistant performs the task using the object managing process. For example, the digital assistant instantiates an object managing process to generate a folder named "Projects" on the desktop of user interface 1310. In some examples, the digital assistant can cause the object managing process to further open the folder either automatically or in response to an additional user input. For example, the digital assistant provides a spoken output 1372 such as "OK, I've created a folder on the desktop called Projects, would you like to open it?" The user provides a speech input 1374 such as "Yes." In response to the user's speech input 1374, the digital assistant causes the object managing process to open the Projects folder and display a user interface 1332 corresponding to the Projects folder.

[0269] With reference to FIG. 13C, in some embodiments, the digital assistant provides one or more affordances that enable the user to manipulate the result of performing the task using the searching process or the object managing process. The one or more affordances include, for example, an edit button, a cancel button, a redo button, an undo button, or the like. For example, as shown in FIG. 13C, after generating the folder named "Projects" on the desktop, the digital assistant provides a user interface 1334, which displays an edit button 1336A, an undo button

1336B, and a redo button 1336C. In some examples, the edit button 1336A enables the user to edit one or more aspects of the object (e.g., edit the name of the Projects folder); the undo button 1336B enables the user to reverse the last task performed by the object managing process (e.g., delete the Projects folder); and the redo button 1336C enables the user to repeat the last task performed by the object managing process (e.g., creating another folder using the object managing process). It is appreciated that the digital assistant can provide any desired affordances to enable the user to perform any manipulation of the result of performing a task using the searching process or the object managing process.

[0270] As described, the digital assistant can determine whether the user intent is to perform a task using a searching process or an object managing process. In some examples, the digital assistant determines that the user intent is not associated with the searching process or the object managing process. For example, the user provides a speech input such as "start dictation." The digital assistant determines that the task of dictation is not associated with searching. In some examples, in accordance with a determination that the task is not associated with searching, the digital assistant further determines whether the task is associated with managing at least one object. For example, the digital assistant determines that the task of dictation is also not associated with managing an object, such as copying, moving, or deleting a file, a folder, or an email. In some examples, in accordance with a determination that the task is not associated with managing an object, the digital assistant determines whether the task can be performed using a process available to the user device. For example, the digital assistant determines that the task of dictation can be performed using a dictation process that is available to the user device. In some examples, the digital assistant initiates a dialog with the user with respect to performing the task using a process available to the user device. For example, the digital assistant provides a spoken output such as "OK, starting dictation." or "Would you like to dictate in this presentation you are working now?" After providing the spoken output, the digital assistant receives a response from the user, for example, confirming that the user intent is to dictate in the presentation the user is currently working on.

5. Exemplary Functions of a Digital Assistant – Continuity

[0271] FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E illustrate functionalities of performing a task at a user device or a first electronic device using remotely located content by a digital assistant. In some examples, the digital assistant system (e.g., digital assistant system 700) is implemented by a user device (e.g., devices 1400, 1500, 1600, and 1700) according to various examples. In some examples, the user device, a server (e.g., server 108), or a combination thereof, may implement a digital assistant system (e.g., digital assistant system 700). The user device can be implemented using, for example, device 104, 200, or 400. In some examples, the user device can be a laptop computer, a desktop computer, or a tablet computer. The user device operates in a multi-tasking environment, such as a desktop environment.

[0272] With references to FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E, in some examples, a user device (e.g., devices 1400, 1500, 1600, and 1700) provides various user interfaces (e.g., user interfaces 1410, 1510, 1610, and 1710). Similar to those described above, the user device displays the various user interfaces on a display, and the various user interfaces enable the user to instantiate one or more processes (e.g., a movie process, a photo process, a web-browsing process).

[0273] As shown in FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E, similar to those described above, the user device (e.g., devices 1400, 1500, 1600, and 1700) displays, on a user interface (e.g., user interfaces 1410, 1510, 1610, and 1710) an affordance (e.g., affordance 1440, 1540, 1640, and 1740) to instantiate a digital assistant service. Similar to those described above, in some examples, the digital assistant is instantiated in response to receiving a pre-determined phrase. In some examples, the digital assistant is instantiated in response to receiving a selection of the affordance.

[0274] With reference to FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E, in some embodiments, a digital assistant receives one or more speech inputs, such as speech inputs 1452, 1454, 1456, 1458, 1552, 1554, 1556, 1652, 1654, 1656, 1752, and 1756 from a user. The user may provide various speech inputs for the purpose of, for example, performing a task at the user device (e.g., devices 1400, 1500, 1600, and 1700) or at a first electronic device (e.g., electronic devices 1420, 1520, 1530, 1522, 1532, 1620, 1622, 1630, 1720, and 1730) using remotely located content. Similar to those described above, in some examples, the digital assistant can

receive speech inputs directly from the user at the user device or indirectly through another electronic device that is communicatively connected to the user device.

[0275] With reference to FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E, in some embodiments, the digital assistant identifies context information associated with the user device. The context information includes, for example, user-specific data, sensor data, and user device configuration data. In some examples, the user-specific data includes log information indicating user preferences, the history of user's interaction with the user device (e.g., devices 1400, 1500, 1600, and 1700), and/or electronic devices communicative connected to the user device, or the like. For example, user-specific data indicates that the user recently took a self-portrait photo using an electronic device 1420 (e.g., a smartphone); that the user recently accessed a podcast, webcast, movie, song, audio book, or the like. In some examples, the sensor data includes various data collected by a sensor associated with the user device or other electronic devices. For example, the sensor data includes GPS location data indicating the physical location of the user device or electronic devices communicatively connected to the user device at any time point or during any time period. For example, the sensor data indicates that a photo stored in electronic device 1420 was taken at Hawaii. In some examples, the user device configuration data includes the current or historical device configurations. For example, the user device configuration data indicates that the user device is currently communicatively connected to some electronic devices but disconnected from other electronic devices. The electronic devices includes, for example, a smartphone, a set-top box, a tablet, or the like. As described in more detail below, the context information can be used in determining a user intent and/or in performing one or more tasks.

[0276] With reference to FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E, similar to those described above, in response to receiving a speech input, the digital assistant determines a user intent based on the speech input. The digital assistant determines the user intent based on a result of natural language processing. For example, the digital assistant identifies an actionable intent based on the user input, and generates a structured query to represent the identified actionable intent. The structured query includes one or more parameters associated with the actionable intent. The one or more parameters can be used to facilitate the performance of a task based on the actionable intent. For example, based on a speech input such as "show the selfie I

just took," the digital assistant determines that the actionable intent is to display a photo, and the parameters include a self-portrait that the user recently took during the past few days. In some embodiments, the digital assistant further determines the user intent based on the speech input and context information. For example, the context information indicates that the user device is communicatively connected to the user's phone using a Bluetooth connection and indicates that a self-portrait photo was added to the user's phone two days ago. As a result, the digital assistant determines that the user intent is to display a photo that is a self-portrait that was added to the user's phone two days ago. Determining the user intent based on speech input and context information is described in more detail below in various examples.

[0277] In some embodiments, in accordance with user intent, the digital assistant further determines whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device. Various examples of the determination are provided below in more detail with respect to FIGs. 14A-14D, 15A-15D, 16A-16C, and 17A-17E.

[0278] With reference to FIG. 14A, in some examples, user device 1400 receives a speech input 1452 from a user to invoke the digital assistant. As shown in FIG. 14A, in some examples, the digital assistant is represented by affordances 1440 or 1441 displayed on user interface 1410. Speech input 1452 includes, for example, "Hey, Assistant." In response to speech input 1452, user device 1400 invokes the digital assistant such that the digital assistant actively monitors subsequent speech inputs. In some examples, the digital assistant provides a spoken output 1472 indicating that it is invoked. For example, spoken output 1472 includes "Go ahead, I am listening." As shown in FIG. 14A, in some examples, user device 1400 is communicatively connected to one or more electronic devices such as electronic device 1420. Electronic device 1420 can communicate with user device 1400 using wired or wireless networks. For example, electronic device 1420 communicates with user device 1400 using Bluetooth connections such that voice and data (e.g., audio and video files) can be exchanged between the two devices.

[0279] With reference to FIG. 14B, in some examples, the digital assistant receives a speech input 1454 such as "Show me the selfie I just took using my phone on this device." Based on speech input 1454 and/or context information, the digital assistant determines the user intent.

For example, as shown in FIG. 14B, context information indicates that the user device 1400 is communicatively connected to electronic device 1420 using wired or wireless networks (e.g., a Bluetooth connection, a Wi-Fi connection, or the like). Context information also indicates that the user recently took a self-portrait, which is stored in electronic device 1420 with a name "selfie0001." As a result, the digital assistant determines that the user intent is to display the photo named selfie0001 stored in electronic device 1420. Alternatively, the photo may have been tagged with photo recognition software as containing the user's face and be identified accordingly.

[0280] As described, in accordance with the user intent, the digital assistant further determines whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device. In some embodiments, determining whether the task is to be performed at the user device or at the first electronic device is based on one or more keywords included in the speech input. For example, the digital assistant determines that speech input 1454 includes keywords or a phrase such as "on this device," indicating the task is to be performed on user device 1400. As a result, the digital assistant determines that displaying the photo named selfie0001 stored in electronic device 1420 is to be performed at user device 1400. User device 1400 and electronic device 1420 are different devices. For example, user device 1400 can be a laptop computer, and electronic device 1420 can be a phone.

[0281] In some embodiments, the digital assistant further determines whether the content associated with the performance of the task is located remotely. Content is located remotely if at or near the time the digital assistant determines which device is to perform the task, at least a portion of the content for performing the task is not stored in the device that is determined to perform the task. For example, as shown in FIG. 14B, at or near the time the digital assistant of user device 1400 determines that the user intent is to display the photo named selfie0001 at user device 1400, the photo named selfie0001 is not stored at user device 1400 and instead is stored at electronic device 1420 (e.g., a smartphone). Accordingly, the digital assistant determines that the photo is located remotely to user device 1400.

[0282] As illustrated in FIG. 14B, in some embodiments, in accordance with a determination that the task is to be performed at the user device and content for performing the task is located

remotely, the digital assistant of the user device receives the content for performing the task. In some examples, the digital assistant of the user device 1400 receives at least a portion of the content stored in the electronic device 1420. For example, to display the photo named selfie0001, the digital assistant of user device 1400 sends a request to electronic device 1420 to obtain the photo named selfie0001. Electronic device 1420 receives the request and, in response, transmits the photo named selfie0001 to user device 1400. The digital assistant of user device 1400 then receives the photo named selfie0001.

[0283] As illustrated in FIG. 14B, in some embodiments, after receiving the remotely located content, the digital assistant provides a response at the user device. In some examples, providing a response includes performing the task using the received content. For example, the digital assistant of user device 1400 displays a user interface 1442 (e.g., a snippet or a window) providing a view 1443 of the photo named selfie0001. View 1443 can be a preview (e.g., a thumbnail), an icon, or a full view of the photo named selfie0001.

[0284] In some examples, providing a response includes providing a link that is associated with the task to be performed at the user device. A link enables instantiating of a process. As described, instantiating a process includes invoking the process if the process is not already running. If at least one instance of the process is running, instantiating a process includes executing an existing instance of the process or generating a new instance of the process. As shown in FIG. 14B, user interface 1442 may provide a link 1444 associated with view 1443 of the photo named selfie0001. Link 1444 enables, for example, instantiating a photo process to view a full representation of the photo or edit the photo. As an example, link 1444 is displayed on the side of view 1443. As another example, view 1443 can itself include or incorporate link 1444 such that a selection of view 1443 instantiates a photo process.

[0285] In some embodiments, providing a response includes providing one or more affordances that enable the user to further manipulate the results of the performance of the task. As shown in FIG. 14B, in some examples, the digital assistant provides affordances 1445 and 1446 on user interface 1442 (e.g., a snippet or a window). Affordance 1445 can include a button for adding a photo to an album, and affordance 1446 can include a button for canceling view 1443 of the photo. The user may select one or both of affordances 1445 and 1446. In response

to the selection of affordance 1445, for example, a photo process adds the photo associated with view 1443 to an album. In response to the selection of affordance 1446, for example, a photo process removes view 1443 from user interface 1442.

[0286] In some embodiments, providing a response includes providing a spoken output according to the task to be performed at the user device. As illustrated in FIG. 14B, the digital assistant represented by affordances 1440 or 1441 provides a spoken output 1474 such as "Here is the last selfie from your phone."

[0287] With reference to FIG. 14C, in some examples, based on a single speech input/utterance and context information, the digital assistant determines that the user intent is to perform a plurality of tasks. As shown in FIG. 14C, the digital assistant receives a speech input 1456 such as "Show me the selfie I just took using my phone on this device and set it as my wallpaper." Based on speech input 1456 and context information, the digital assistant determines that the user intent is to perform a first task of displaying the photo named selfie0001 stored at electronic device 1420 and performs a second task of setting the photo named selfie0001 as the wallpaper. Thus, based on a single speech input 1456, the digital assistant determines that the user intent is to perform multiple tasks.

[0288] In some embodiments, the digital assistant determines whether the plurality of tasks is to be performed at the user device or at an electronic device communicatively connected to the user device. For example, using the keywords "this device" included in speech input 1456, the digital assistant determines that the plurality of tasks is to be performed at user device 1400. Similar to those described above, the digital assistant further determines whether the content for performing at least one task is located remotely. For example, the digital assistant determines that the content for performing at least the first task (e.g., displaying the photo named selfie0001) is located remotely. In some embodiments, in accordance with a determination that the plurality of tasks is to be performed at the user device and content for performing at least one task is located remotely, the digital assistant requests the content from another electronic device (e.g., electronic device 1420), receives the content for performing the tasks, and provides a response at the user device.

[0289] In some embodiments, providing a response includes performing the plurality of tasks. For example, as illustrated in FIG. 14C, providing a response includes performing the first task of displaying a view 1449 of the photo named selfie0001, and performing the second task of setting the photo named selfie0001 as the wallpaper. In some examples, the digital assistant automatically configures the wallpaper to be the photo named selfi0001 using a desktop settings configuration process. In some examples, the digital assistant provides a link to desktop settings 1450, enabling the user to manually configure the wallpaper using the photo named selfie0001. For example, the user may select the link to desktop settings 1450 by using an input device such as a mouse, a stylus, or a finger. Upon receiving the selection of the link to desktop setting 1450, the digital assistant initiates the desktop setting configuration process that enables the user to select the photo named selfie0001 and set it as the wallpaper of user device 1400.

[0290] As illustrated in FIG. 14C, in some examples, the digital assistant initiates a dialog with the user and facilitates the configuration of the wallpaper in response to receiving a speech input from the user. For example, the digital assistant provides a spoken output 1476 such as "Here is the last selfie from your phone. Set is as wallpaper?" The user provides a speech input such as "OK." Upon receiving the speech input, the digital assistant instantiates the desktop settings configuration process to configure the wallpaper as the photo named selfie0001.

[0291] As described, in some examples, the digital assistant determines the user intent based on the speech input and context information. With reference to FIG. 14D, in some examples, the speech input may not include information sufficient to determine the user intent. For example, the speech input may not indicate the location of the content for performing the task. As shown in FIG. 14D, the digital assistant receives a speech input 1458 such as "Show me the selfie I just took." Speech input 1458 does not include one or more keywords indicating which photo is to be displayed or the location of the selfie to be displayed. As a result, the user intent may not be determined based solely on speech input 1458. In some examples, the digital assistant determines the user intent based on speech input 1458 and context information. For example, based on context information, the digital assistant determines that user device 1400 is communicatively connected to electronic device 1420. In some examples, the digital assistant instantiates a searching process to search for photos that the user recently took at user device 1400 and electronic device 1420. Based on the search result, the digital assistant determines that

a photo named selfie0001 is stored in electronic device 1420. Accordingly, the digital assistant determines that the user intent is to display the photo named selfie0001 located at electronic device 1420. In some examples, if the user intent cannot be determined based on the speech input and context information, the digital assistant initiates a dialog with the user to further clarify or disambiguate the user intent.

[0292] As illustrated in FIG. 14D, in some examples, the speech input may not include one or more keywords indicating whether a task is to be performed at the user device or at an electronic device communicatively connected to the user device. For example, speech input 1458 does not indicate whether the task of displaying the selfie is to be performed at user device 1400 or at electronic device 1420. In some examples, the digital assistant determines whether a task is to be performed at the user device or at an electronic device based on context information. As an example, the context information indicates that the digital assistant receives speech input 1458 at user device 1400, not at electronic device 1420. As a result, the digital assistant determines that the task of displaying the selfie is to be performed at user device 1400. As another example, context information indicates that a photo is to be displayed on electronic device 1420 according to user preferences. As a result, the digital assistant determines that the task of displaying the selfie is to be performed at electronic device 1420. It is appreciated that the digital assistant can determine whether a task is to be performed at the user device or at an electronic device based on any context information.

[0293] With reference to FIG. 15A, in some embodiments, a digital assistant determines that the task is to be performed at an electronic device (e.g., electronic device 1520 and/or 1530) communicatively connected to the user device (e.g., user device 1500) and determine that the content is located remotely to the electronic device. As shown in FIG. 15A, in some examples, the digital assistant receives a speech input 1552 such as “Play this movie on my TV.” As described, the digital assistant can determine the user intent based on speech input 1552 and context information. For example, context information indicates that user interface 1542 is displaying a movie named ABC.mov. As a result, the digital assistant determines that the user intent is to play the movie named ABC.mov.

[0294] In accordance with the user intent, the digital assistant furthers determine whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device. In some embodiments, determining whether the task is to be performed at the user device or at the first electronic device is based on one or more keywords included in the speech input. For example, speech input 1552 includes the words or phrase "on my TV." In some examples, context information indicates that user device 1500 is connected to a set-top box 1520 and/or a TV 1530 using, for example, a wired connection, a Bluetooth connection, or a Wi-Fi connection. As a result, the digital assistant determines that the task of playing the movie named ABC.mov is to be performed on set-top box 1520 and/or TV 1530.

[0295] In some embodiments, the digital assistant further determines whether the content associated with the performance of the task is located remotely. As described, content is located remotely if at or near the time the digital assistant determines which device is to perform the task, at least a portion of the content for performing the task is not stored in the device that is determined to perform the task. For example, as shown in FIG. 15A, at or near the time the digital assistant of user device 1500 determines that movie ABC.mov is to be played at set-top box 1520 and/or TV1530, at least a portion of the movie ABC.mov is stored at user device 1500 (e.g., a laptop computer) and/or a server (not shown) and is not stored at set-top box 1520 and/or TV 1530. Accordingly, the digital assistant determines that the movie ABC.mov is located remotely to set-top box 1520 and/or TV 1530.

[0296] With reference to FIG. 15B, in accordance with a determination that the task is to be performed at the first electronic device (e.g., set-top box 1520 and/or TV 1530) and the content for performing the task is located remotely to the first electronic device, the digital assistant of the user device provides the content to the first electronic device to perform the task. For example, to play the movie ABC.mov on set-top box 1520 and/or TV 1530, the digital assistant of user device 1500 transmits at least a portion of the movie ABC.mov to set-top box 1520 and/or TV 1530.

[0297] In some examples, instead of providing the content from the user device, the digital assistant of the user device causes at least a portion of the content to be provided from another electronic device (e.g., a server) to the first electronic device to perform the task. For example,

the movie ABC.mov is stored in a server (not shown) and not at user device 1500. As a result, the digital assistant of user device 1500 causes at least a portion of the movie named ABC.mov to be transmitted from the server to set-top box 1520 and/or TV 1530. In some examples, the content for performing the task is provided to set-top box 1520, which then transmits the content to TV 1530. In some examples, the content for performing the task is provided to TV 1530 directly.

[0298] As illustrated in FIG. 15B, in some examples, after the content is provided to the first electronic device (e.g., set-top box 1520 and/or TV 1530), the digital assistant of user device 1500 provides a response at user device 1500. In some examples, providing the response includes causing the task to be performed at set-top box 1520 and/or TV 1530 using the content. For example, the digital assistant of user device 1500 sends a request to set-top box 1520 and/or TV 1530 to initiate a multimedia process to play the movie ABC.mov. In response to the request, set-top box 1520 and/or TV 1530 initiates the multimedia process to play the movie ABC.mov.

[0299] In some examples, the task to be performed at the first electronic device (e.g., set-top box 1520 and/or TV 1530) is a continuation of a task performed remotely to the first electronic device. For example, as illustrated in FIGs. 15A and 15B, the digital assistant of user device 1500 has caused a multimedia process of user device 1500 to play a portion of the movie ABC.mov at user device 1500. In accordance with the determination that the user intent is to play the movie ABC.mov at the first electronic device (e.g., set-top box 1520 and/or TV 1530), the digital assistant of user device 1500 causes the first electronic device to continue playing the rest of the movie ABC.mov rather than start playing from the beginning. As a result, the digital assistant of user device 1500 enables the user to continuously watch the movie.

[0300] As illustrated in FIG. 15B, in some embodiments, providing a response includes providing one or more affordances that enable the user to further manipulate the results of the performance of the task. As shown in FIG. 15B, in some examples, the digital assistant provides affordances 1547 and 1548 on a user interface 1544 (e.g., a snippet or a window). Affordance 1547 can be a button for cancelling the playing of movie ABC.mov on the first electronic device (e.g., set-top box 1520 and/or TV 1530). Affordance 1548 can be a button to pause or resume

the playing of movie ABC.mov that is playing on the first electronic device. The user may select affordance 1547 or 1548 using an input device such as a mouse, a stylus, or a finger. Upon receiving a selection of affordance 1547, for example, the digital assistant causes the playing of movie ABC.mov on the first electronic device to stop. In some examples, after the playing on the first electronic device stops, the digital assistant also causes the playing of movie ABC.mov on user device 1500 to resume. Upon receiving a selection of affordance 1548, for example, the digital assistant causes the playing of movie ABC.mov on the first electronic device to pause or resume.

[0301] In some embodiments, providing a response includes providing a spoken output according to the task to be performed at the first electronic device. As illustrated in FIG. 15B, the digital assistant represented by affordance 1540 or 1541 provides a spoken output 1572 such as "Playing your movie on TV."

[0302] As described, in accordance with a determination that the task is to be performed at a first electronic device and the content for performing the task is located remotely to the first electronic device, the digital assistant provides the content for performing the task to the first electronic device. With reference to FIG. 15C, the content for performing the task can include, for example, a document (e.g., document 1560) or location information. For instance, the digital assistant of user device 1500 receives a speech input 1556 such as "Open this pdf on my tablet." The digital assistant determines that the user intent is to perform a task of displaying document 1560 and determines that the task is to be performed at a tablet 1532 that is communicatively connected to user device 1500. As a result, the digital assistant provides document 1560 to tablet 1532 to be displayed. As another example, the digital assistant of user device 1500 receives a speech input 1554 such as "Send this location to my phone." The digital assistant determines that the user intent is to perform a task of navigation using the location information and determines that the task is to be performed at phone 1522 (e.g., a smartphone) that is communicatively connected to user device 1500. As a result, the digital assistant provides location information (e.g., 1234 Main St.) to phone 1522 to perform the task of navigation.

[0303] As described, in some examples, after providing the content for performing the task to the first electronic device, the digital assistant provides a response at the user device. In some

embodiments, providing the response includes causing the task to be performed at the first electronic device. For example, as shown in FIG. 15D, the digital assistant of user device 1500 transmits a request to phone 1522 to perform the task of navigating to the location 1234 Main St. The digital assistant of user device 1500 further transmits a request to tablet 1532 to perform the task of displaying document 1560. In some examples, providing the response at the user device includes providing a spoken output according to the task to be performed at the first electronic device. As illustrated in FIG. 15D, the digital assistant provides a spoken output 1574 such as "Showing the pdf on your tablet" and a spoken output 1576 such as "navigating to 1234 Main St on your phone."

[0304] As described, in some examples, the speech input may not include one or more keywords indicating whether a task is to be performed at the user device or at a first electronic device communicatively connected to the user device. With reference to FIG. 16A, for example, the digital assistant receives a speech input 1652 such as "Play this movie." Speech input 1652 does not indicate whether the task of playing the movie is to be performed at user device 1600 or at a first electronic device (e.g., set-top box 1620 and/or TV 1630, phone 1622, or tablet 1632).

[0305] In some embodiments, to determine whether the task is to be performed at the user device or at a first electronic device, the digital assistant of the user device determines whether performing the task at the user device satisfies performance criteria. Performance criteria facilitate evaluating the performance of the task. For example, as illustrated in FIG. 16A, the digital assistant determines that the user intent is to perform the tasking of playing the movie ABC.mov. Performance criteria for playing a movie include, for example, the quality criteria of playing a movie (e.g., 480p, 720p, 1080p), the smoothness criteria of playing the movie (e.g., no delay or waiting), the screen size criteria (e.g., a minimum screen size of 48 inches), the sound effect criteria (e.g., stereo sounds, number of speakers), or the like. The performance criteria can be pre-configured and/or dynamically updated. In some examples, the performance criteria are determined based on context information such as user-specific data (e.g., user preferences), device configuration data (e.g., screen resolution and size of the electronic devices), or the like.

[0306] In some examples, the digital assistant of user device 1600 determines that performing the task at the user device satisfies the performance criteria. For example, as

illustrated in FIG. 16A, user device 1600 may have a screen resolution, a screen size, and sound effect that satisfy the performance criteria of playing the movie ABC.mov, which may be a low-resolution online video. In accordance with a determination that performing the task at user device 1600 satisfies the performance criteria, the digital assistant determines that the task is to be performed at user device 1600.

[0307] In some examples, the digital assistant of user device 1600 determines that performing the task at the user device does not satisfy the performance criteria. For example, user device 1600 may not have the screen size, the resolution, and/or the sound effect to satisfy the performance criteria of playing the movie ABC.mov, which may be a high-resolution Blu-ray video. In some examples, in accordance with a determination that performing the task at the user device does not satisfy the performance criteria, the digital assistant of user device 1600 determines whether performing the task at the first electronic device satisfies the performance criteria. As illustrated in FIG. 16B, the digital assistant of user device 1600 determines that performing the task of playing the movie ABC.mov at set-top box 1620 and/or TV 1630 satisfies the performance criteria. For example, set-top box 1620 and/or TV 1630 may have a screen size of 52 inches, may have a 1080p resolution, and may have eight speakers connected. As a result, the digital assistant determines that the task is to be performed at set-top box 1620 and/or TV 1630.

[0308] In some examples, the digital assistant of user device 1600 determines that performing the task at the first electronic device does not satisfy the performance criteria. In accordance with the determination, the digital assistant determines whether performing the task at the second electronic device satisfies the performance criteria. For example, as illustrated in FIG. 16B, TV 1630 may have a screen resolution (e.g., 720p) that does not satisfy the performance criteria (e.g., 1080p). As a result, the digital assistant determines whether any one of phone 1622 (e.g., a smartphone) or tablet 1632 satisfies the performance criteria.

[0309] In some examples, the digital assistant determines which device provides the optimum performance of the task. For example, as illustrated in FIG. 16B, the digital assistant evaluates or estimates the performance of the task of playing movie ABC.mov on each of user device 1600, set-top box 1620 and TV 1630, phone 1622, and tablet 1632. Based on the

evaluation or estimation, the digital assistant determines whether performing the task at one device (e.g., user device 1600) is better than at another device (e.g., phone 1622) and determines a device for optimum performance.

[0310] As described, in some examples, in accordance with the determination of a device for performing the task, the digital assistant provides a response at user device 1600. In some embodiments, providing a response includes providing a spoken output according to the task to be performed at the device. As illustrated in FIG. 16B, the digital assistant represented by affordances 1640 or 1641 provides a spoken output 1672 such as "I will play this movie on your TV, proceed?" In some examples, the digital assistant receives a speech input 1654 such as "OK" from the user. In response, the digital assistant causes the movie ABC.mov to be played at, for example, set-top box 1620 and TV 1630 and provides a spoken output 1674 such as "Playing your movie on your TV."

[0311] In some examples, providing a response includes providing one or more affordances that enable the user to select another electronic device for performance of the task. As illustrated in FIG. 16B, for example, the digital assistant provides affordances 1655A-B (e.g., a cancel button and a tablet button). Affordance 1655A enables the user to cancel playing the movie ABC.mov at set-top box 1620 and TV 1630. Affordance 1655B enables the user to select tablet 1632 to continue playing the movie ABC.mov.

[0312] With reference to FIG. 16C, in some embodiments, to determine a device for performing a task, the digital assistant of user device 1600 initiates a dialog with the user. For example, the digital assistant provides a spoken output 1676 such as "Should I play your movie on the TV or on the tablet?" The user provides a speech input 1656 such as "On my tablet." Upon receiving speech input 1656, the digital assistant determines that the task of playing the movie is to be performed at tablet 1632, which is communicatively connected to user device 1600. In some examples, the digital assistant further provides a spoken output 1678 such as "Playing your movie on your tablet."

[0313] With reference to FIG. 17A, in some embodiments, a digital assistant of a user device 1700 continues to perform a task that was partially performed remotely at a first electronic device. In some embodiments, the digital assistant of a user device continues to perform the task

using content received from a third electronic device. As illustrated in FIG. 17A, in some examples, phone 1720 may have been performing a task of flight booking using content from a third electronic device such as a server 1730. For example, the user may have been using phone 1720 to book flights from Kayak.com. As a result, phone 1720 receives content transmitted from server 1730 that is associated with Kayak.com. In some examples, the user may be interrupted while booking his or her flight on phone 1720 and may desire to continue the flight booking using user device 1700. In some examples, the user may desire to continue the flight booking simply because using user device 1700 is more convenient. Accordingly, the user may provide a speech input 1752 such as "Continue the flight booking on Kayak from my phone."

[0314] With reference to FIG. 17B, upon receiving speech input 1752, the digital assistant determines the user intent is to perform a task of flight booking. In some examples, the digital assistant further determines that the task is to be performed at user device 1700 based on context information. For example, the digital assistant determines that speech input 1752 is received at user device 1700 and therefore determines that the task is to be performed at user device 1700. In some examples, the digital assistant further uses context information such as user preferences (e.g., user device 1700 is used frequently in the past for flight booking) to determine that the task is to be performed at user device 1700.

[0315] As shown in FIG. 17B, in accordance with the determination that the task is to be performed at the user device 1700, and the content for performing the task is located remotely, the digital assistant receives the content for performing the task. In some examples, the digital assistant receives the at least a portion of the content from phone 1720 (e.g., a smartphone) and/or at least a portion of the content from server 1730. For example, the digital assistant receives data representing the status of flight booking from phone 1720 such that user device 1700 can continue the flight booking. In some examples, the data representing the status of flight booking is stored at server 1730, such as a server associated with Kayak.com. The digital assistant thus receives data from server 1730 for continuing the flight booking.

[0316] As illustrated in FIG. 17B, after receiving the content from phone 1720 and/or server 1730, the digital assistant provides a response at user device 1700. In some examples, providing the response includes continuing to perform the task of flight booking that was partially

performed remotely at phone 1720. For example, the digital assistant displays a user interface 1742 enabling the user to continue booking the flight on Kayak.com. In some examples, providing the response includes providing a link associated with the task to be performed at user device 1700. For example, the digital assistant displays a user interface 1742 (e.g., a snippet or a window) providing the current status of flight booking (e.g., showing available flights). User interface 1742 also provides a link 1744 (e.g., a link to a web browser) for continuing performing the task of flight booking. In some embodiments, the digital assistant also provides a spoken output 1772 such as "Here is the booking on Kayak. Continue in your web browser?"

[0317] As shown in FIGs. 17B and 17C, for example, if the user selects link 1744, the digital assistant instantiates a web browsing process and displays a user interface 1746 (e.g., a snippet or a window) for continuing the flight booking task. In some examples, in response to spoken output 1772, the user provides a speech input 1756 such as "OK" confirming that the user desires to continue flight book using a web browser of user device 1700. Upon receiving speech input 1756, the digital assistant instantiates a web browsing process and displays user interface 1746 (e.g., a snippet or a window) for continuing the flight booking task.

[0318] With reference to FIG. 17D, in some embodiments, a digital assistant of a user device 1700 continues to perform a task that was partially performed remotely at a first electronic device. In some embodiments, the digital assistant of the user device continues to perform the task using content received from the first electronic device, rather than a third electronic device such as a server. As illustrated in FIG. 17D, in some examples, the first electronic device (e.g., phone 1720 or tablet 1732) may have been performing a task. For example, the user may have been using phone 1720 to compose an email or using tablet 1732 to edit a document such as a photo. In some examples, the user is interrupted while using phone 1720 or tablet 1732, and/or desires to continue the performance of the task using user device 1700. In some examples, the user may desire to continue the performance of the task simply because using user device 1700 is more convenient (e.g., a larger screen). Accordingly, the user may provide a speech input 1758 such as "Open the document I was just editing" or speech input 1759 such as "Open the email I was just drafting."

[0319] With reference to FIG. 17D, upon receiving speech input 1758 or 1759, the digital assistant determines the user intent is to perform a task of editing a document or composing an email. Similar to those described above, in some examples, the digital assistant further determines that the task is to be performed at user device 1700 based on context information, and determines that the content for performing the task is located remotely. Similar to described above, in some examples, the digital assistant determines, based on context information (e.g., user-specific data), that the content is located remotely at the first electronic device (e.g., at phone 1720 or tablet 1732), rather than at a server. As shown in FIG. 17D, in accordance with the determination that the task is to be performed at the user device 1700 and the content for performing the task is located remotely, the digital assistant receives the content for performing the task. In some examples, the digital assistant receives the at least a portion of the content from phone 1720 (e.g., a smartphone) and/or at least a portion of the content from tablet 1730. After receiving the content from phone 1720 and/or tablet 1732, the digital assistant provides a response at user device 1700, such as displaying a user interface 1748 for the user to continue editing the document and/or displaying a user interface 1749 for the user to continue composing the email. It is appreciated that the digital assistant of user device 1700 can also cause a first electronic device to continue performing a task that was partially performed remotely at the user device 1700. For example, the user may be composing an email on user device 1700 and may need to leave. The user provides a speech input such as "Open the email I was drafting on my phone." Based on the speech input, the digital assistant determines the user intent is to continue performing the task on phone 1720 and the content is located remotely at the user device 1700. In some examples, the digital assistant provides the content for performing the task to the first electronic device and causes the first electronic device to continue performing the task, similar to those described above.

[0320] With reference to FIG. 17E, in some embodiments, continuing to performing a task is based on context information that is shared or synchronized among a plurality of devices including, for example, user device 1700 and first electronic device (e.g., phone 1720). As described, in some examples, the digital assistant determines a user intent based on the speech input and context information. The context information can be stored locally or remotely. For example, as shown in FIG. 17E, the user provides a speech input 1760 such as "What is the weather like in New York?" to phone 1720. A digital assistant of phone 1720 determines the

user intent, performs the task to obtain the weather information in New York, and displays the weather information of New York on a user interface of phone 1720. The user subsequently provides a speech input 1761 such as "How about in Los Angeles?" to user device 1700. In some examples, the digital assistant of user device 1700 determines the user intent using context information stored at and/or shared by phone 1720, either directly or through a server. The context information includes, for example, historical user data associated with phone 1720, conversational state, system state, etc. Both the historical user data and conversational state indicate that user was inquiring about weather information. Accordingly, the digital assistant of user device 1700 determines that the user intent is to obtain the weather information in Los Angeles. Based on the user intent, the digital assistant of user device 1700 receives the weather information from, for example, a server, and provides a user interface 1751 displaying the weather information on user device 1710.

6. Exemplary Functions of a Digital Assistant – Voice-enabled System Configuration Management

[0321] FIGs. 18A-18F and 19A-19D illustrate functionalities of providing system configuration information or performing a task in response to a user request by a digital assistant. In some examples, the digital assistant system (e.g., digital assistant system 700) can be implemented by a user device according to various examples. In some examples, the user device, a server (e.g., server 108), or a combination thereof, may implement a digital assistant system (e.g., digital assistant system 700). The user device is implemented using, for example, device 104, 200, or 400. In some examples, the user device is a laptop computer, a desktop computer, or a tablet computer. The user device operates in a multi-tasking environment, such as a desktop environment.

[0322] With references to FIGs. 18A-18F and 19A-19D, in some examples, a user device provides various user interfaces (e.g., user interfaces1810 and 1910). Similar to those described above, the user device displays the various user interfaces on a display and the various user interfaces enable the user to instantiate one or more processes (e.g., system configuration processes).

[0323] As shown in FIGs. 18A-18F and 19A-19D, similar to those described above, the user device displays, on a user interface (e.g., user interfaces 1810 and 1910), an affordance (e.g., affordance 1840 and 1940) to facilitate the instantiation of a digital assistant service.

[0324] Similar to those described above, in some examples, the digital assistant is instantiated in response to receiving a pre-determined phrase. In some examples, the digital assistant is instantiated in response to receiving a selection of the affordance.

[0325] With reference to FIGs. 18A-18F and 19A-19D, in some embodiments, a digital assistant receives one or more speech inputs, such as speech inputs 1852, 1854, 1856, 1858, 1860, 1862, 1952, 1954, 1956, and 1958 from a user. The user provides various speech inputs for the purpose of managing one or more system configurations of the user device. The system configurations can include audio configurations, date and time configurations, dictation configuration, display configurations, input device configurations, notification configurations, printing configurations, security configurations, backup configurations, application configurations, user interface configurations, or the like. To manage audio configurations, a speech input may include "Mute my microphone," "Turn the volume all the up," "Turn the volume up 10%," or the like. To manage date and time configurations, a speech input may include "What is my time zone?", "Change my time zone to Cupertino Time," "Add a clock for London time zone," or the like. To manage dictation configurations, a speech input may include "Turn on dictation," "Turn off dictation," "Dictation in Chinese," "Enable advanced commands," or the like. To manage display configurations, a speech input may include "Make my screen brighter," "Increase the contrast my 20%," "Extend my screen to a second monitor," "Mirror my display," or the like. To manage input device configurations, a speech input may include "Connect my Bluetooth keyboard," "Make my mouse pointer bigger," or the like. To manage network configurations, a speech input may include "Turn Wi-Fi on," "Turn Wi-Fi off," "Which Wi-Fi network am I connected to?", "Am I connected to my phone?", or the like. To manage notification configuration, a speech input may include "Turn on Do not Disturb," "Stop showing me these notifications," "Show only new emails," "No alert for text message," or the like. To manage printing configurations, a speech input may include "Does my printer have enough ink?", "Is my printer connected?", or the like. To manage security configurations, a speech input may include "Change password for John's account," "Turn on firewall," "Disable cookie," or the

like. To manage backup configurations, a speech input may include "Run backup now," "Set backup interval to once a month," "Recover the July 4 backup of last year," or the like. To manage application configurations, a speech input may include "Change my default web browser to Safari," "Automatically log in to Messages application each time I sign in," or the like. To manage user interface configurations, a speech input may include "Change my desktop wallpapers," "Hide the dock," "Add Evernote to the Dock," or the like. Various examples of using speech inputs to manage system configurations are described below in more details.

[0326] Similar to those described above, in some examples, the digital assistant receives speech inputs directly from the user at the user device or indirectly through another electronic device that is communicatively connected to the user device.

[0327] With reference to FIGs. 18A-18F and 19A-19D, in some embodiments, the digital assistant identifies context information associated with the user device. The context information includes, for example, user-specific data, sensor data, and user device configuration data. In some examples, the user-specific data includes log information indicating user preferences, the history of user's interaction with the user device, or the like. For example, user-specific data indicates the last time the user's system was backed up; and that the user's preferences of a particular Wi-Fi network when several Wi-Fi networks are available or the like. In some examples, the sensor data includes various data collected by a sensor. For example, the sensor data indicates a printer ink level collected by a printer ink level sensor. In some examples, the user device configuration data includes the current and historical device configurations. For example, the user device configuration data indicates that the user device is currently communicatively connected to one or more electronic devices using Bluetooth connections. The electronic devices may include, for example, a smartphone, a set-top box, a tablet, or the like. As described in more detail below, the user device can determine user intent and/or perform one or more processes using the context information.

[0328] With reference to FIGs. 18A-18F and 19A-19D, similar to those described above, in response to receiving a speech input, the digital assistant determines a user intent based on the speech input. The digital assistant determines the user intent based on a result of natural language processing. For example, the digital assistant identifies an actionable intent based on

the user input, and generates a structured query to represent the identified actionable intent. The structured query includes one or more parameters associated with the actionable intent. The one or more parameters can be used to facilitate the performance of a task based on the actionable intent. For example, based on a speech input such as "Turn the volume up by 10%," the digital assistant determines that the actionable intent is to adjust the system volume, and the parameters include setting the volume to be 10% higher than the current volume level. In some embodiments, the digital assistant also determines the user intent based on the speech input and context information. For example, the context information may indicate that the current volume of the user device is at 50%. As a result, upon receiving the speech input such as "Turn the volume up by 10%," the digital assistant determines that the user intent is to increase the volume level to 60%. Determining the user intent based on speech input and context information is described in more detail below in various examples.

[0329] In some embodiments, the digital assistant further determines whether the user intent indicates an informational request or a request for performing a task. Various examples of the determination are provided below in more detail with respect to FIGs. 18A-18F and 19A-19D.

[0330] With reference to FIG. 18A, in some examples, the user device displays a user interface 1832 associated with performing a task. For example, the task includes composing a meeting invitation. In composing the meeting invitation, the user may desire to know the time zone of the user device so that the meeting invitation can be properly composed. In some examples, the user provides a speech input 1852 to invoke the digital assistant represented by affordance 1840 or 1841. Speech input 1852 includes, for example, "Hey, Assistant." The user device receives the speech input 1852 and, in response, invokes the digital assistant such that the digital assistant actively monitors subsequent speech inputs. In some examples, the digital assistant provides a spoken output 1872 indicating that it is invoked. For example, spoken output 1872 includes "Go ahead, I am listening."

[0331] With reference to FIG. 18B, in some examples, the user provides a speech input 1854 such as "What is my time zone?" The digital assistant determines that the user intent is to obtain the time zone of the user device. The digital assistant further determines whether the user intent indicates an informational request or a request for performing a task. In some examples,

determining whether the user intent indicates an informational request or a request for performing a task includes determining whether the user intent is to vary a system configuration. For example, based on the determination that the user intent is to obtain the time zone of the user device, the digital assistant determines that no system configuration is to be varied. As a result, the digital assistant determines that the user intent indicates an informational request.

[0332] In some embodiments, in accordance with a determination that the user intent indicates an informational request, the digital assistant provides a spoken response to the informational request. In some examples, the digital assistant obtains status of one or more system configurations according to the informational request, and provides the spoken response according to the status of one or more system configurations. As shown in FIG. 18B, the digital assistant determines that the user intent is to obtain the time zone of the user device, and this user intent indicates an informational request. Accordingly, the digital assistant obtains the time zone status from the time and date configuration of the user device. The time zone status indicates, for example, the user device is set to the Pacific time zone. Based on the time zone status, the digital assistant provides a spoken output 1874 such as "Your computer is set to Pacific Standard Time." In some examples, the digital assistant further provides a link associated with the informational request. As illustrated in FIG. 18B, the digital assistant provides a link 1834, enabling the user to further manage the data and time configurations. In some examples, the user uses an input device (e.g., a mouse) to select link 1834. Upon receiving the user's selection of link 1834, the digital assistant instantiates a date and time configuration process and displays an associated date and time configuration user interface. The user can thus use the date and time configuration user interface to further manage the date and time configurations.

[0333] With reference to FIG. 18C, in some examples, the user device displays a user interface 1836 associated with performing a task. For example, the task includes playing a video (e.g., ABC.mov). To enhance the experience of watching the video, the user may desire to use a speaker and may want to know whether a Bluetooth speaker is connected. In some examples, the user provides a speech input 1856 such as "Is my Bluetooth speaker connected?" The digital assistant determines that the user intent is to obtain the connection status of the Bluetooth speaker 1820. The digital assistant further determines that obtaining the connection status of the

Bluetooth speaker 1820 does not vary any system configuration and therefore is an informational request.

[0334] In some embodiments, in accordance with a determination that the user intent indicates an informational request, the digital assistant obtains status of system configurations according to the informational request, and provides the spoken response according to the status of the system configurations. As shown in FIG. 18C, the digital assistant obtains the connection status from the network configuration of the user device. The connection status indicates, for example, user device 1800 is not connected to a Bluetooth speaker 1820. Based on the connection status, the digital assistant provides a spoken output 1876 such as "No, it is not connected, you can check Bluetooth devices in the network configurations." In some examples, the digital assistant further provides a link associated with the informational request. As illustrated in FIG. 18C, the digital assistant provides a link 1838, enabling the user to further manage the network configurations. In some examples, the user uses an input device (e.g., a mouse) to select link 1838. Upon receiving the user's selection of link 1838, the digital assistant instantiates a network configuration process and displays an associated network configuration user interface. The user can thus use the network configuration user interface to further manage the network configurations.

[0335] With reference to FIG. 18D, in some examples, the user device displays a user interface 1842 associated with performing a task. For example, the task includes viewing and/or editing a document. The user may desire to print out the document and may want to know whether a printer 1830 has enough ink for the printing job. In some examples, the user provides a speech input 1858 such as "Does my printer have enough ink?" The digital assistant determines that the user intent is to obtain printer ink level status of the printer. The digital assistant further determines that the obtaining the printer level status does not vary any system configuration and therefore is an informational request.

[0336] In some embodiments, in accordance with a determination that the user intent indicates an informational request, the digital assistant obtains status of system configurations according to the informational request, and provides the spoken response according to the status of the system configurations. As shown in FIG. 18D, the digital assistant obtains the printer ink

level status from the printing configuration of the user device. The printer ink level status indicates, for example, the printer ink level of printer 1830 is at 50%. Based on the connection status, the digital assistant provides a spoken output 1878 such as "Yes, your printer has enough ink. You can also look up printer supply levels in the printer configurations." In some examples, the digital assistant further provides a link associated with the informational request. As illustrated in FIG. 18D, the digital assistant provides a link 1844, enabling the user to further manage the printer configurations. In some examples, the user uses an input device (e.g., a mouse) to select link 1844. Upon receiving the user's selection of the link, the digital assistant instantiates a printer configuration process and displays an associated printer configuration user interface. The user can thus use the printer configuration user interface to further manage the printer configurations.

[0337] With reference to FIG. 18E, in some examples, the user device displays a user interface 1846 associated with performing a task. For example, the task includes browsing Internet using a web browser (e.g., Safari). To browse the Internet, the user may desire to know available Wi-Fi networks and select one Wi-Fi network to connect. In some examples, the user provides a speech input 1860 such as "Which Wi-Fi networks are available?" The digital assistant determines that the user intent is to obtain a list of available Wi-Fi networks. The digital assistant further determines that obtaining the list of available Wi-Fi networks does not vary any system configuration and therefore is an informational request.

[0338] In some embodiments, in accordance with a determination that the user intent indicates an informational request, the digital assistant obtains status of system configurations according to the informational request, and provides the spoken response according to the status of the system configurations. As shown in FIG. 18E, the digital assistant obtains status of currently available Wi-Fi networks from the network configuration of the user device. The status of currently available Wi-Fi networks indicates, for example, Wi-Fi network 1, Wi-Fi network 2, and Wi-Fi network 3 are available. In some examples, the status further indicates the signal strength of each of the Wi-Fi networks. The digital assistant displays a user interface 1845 providing information according to the status. For example, user interface 1845 provides the list of available Wi-Fi networks. The digital assistant also provides a spoken output 1880 such as "Here is a list of available Wi-Fi networks." In some examples, the digital assistant

further provides a link associated with the informational request. As illustrated in FIG. 18E, the digital assistant provides a link 1847, enabling the user to further manage the network configurations. In some examples, the user uses an input device (e.g., a mouse) to select link 1847. Upon receiving the user's selection of the link 1847, the digital assistant instantiates a network configuration process and displays an associated network configuration user interface. The user can thus use the network configuration user interface to further manage the configurations.

[0339] With reference to FIG. 18F, in some examples, the user device displays a user interface 1890 associated with performing a task. For example, the task includes preparing a meeting agenda. In preparing a meeting agenda, the user may desire to find a date and time for the meeting. In some examples, the user provides a speech input 1862 such as "Find a time on my calendar for next Tuesday's meeting in the morning." The digital assistant determines that the user intent is to find an available time slot on the user's calendar on Tuesday morning. The digital assistant further determines that finding a time slot does not vary any system configuration and therefore is an informational request.

[0340] In some embodiments, in accordance with a determination that the user intent indicates an informational request, the digital assistant obtains status of system configurations according to the informational request, and provides the spoken response according to the status of the system configurations. As shown in FIG. 18F, the digital assistant obtains status of user's calendar from calendar configurations. The status of user's calendar indicates, for example, 9a.m. or 11a.m. on Tuesday is still available. The digital assistant displays a user interface 1891 providing information according to the status. For example, user interface 1891 provides the user's calendar in the proximity of the date and time the user requested. In some examples, the digital assistant also provides a spoken output 1882 such as "It looks like Tuesday 9a.m. or 11a.m is available." In some examples, the digital assistant further provides a link associated with the informational request. As illustrated in FIG. 18F, the digital assistant provides a link 1849, enabling the user to further manage the calendar configurations. In some examples, the user uses an input device (e.g., a mouse) to select link 1849. Upon receiving the user's selection of link 1849, the digital assistant instantiates a calendar configuration process and displays an

associated calendar configuration user interface. The user can thus use the calendar configuration user interface to further manage the configurations.

[0341] With reference to FIG. 19A, the user device displays a user interface 1932 associated with performing a task. For example, the task includes playing a video (e.g., ABC.mov). While the video is playing, the user may desire to turn up the volume. In some examples, the user provides a speech input 1952 such as "Turn the volume all the way up." The digital assistant determines that the user intent is to increase the volume to its maximum level. The digital assistant further determines whether the user intent indicates an informational request or a request for performing a task. For example, based on the determination that the user intent is to increase the volume of the user device, the digital assistant determines that an audio configuration is to be varied, and therefore the user intent indicates a request for performing a task.

[0342] In some embodiments, in accordance with a determination that the user intent indicates a request for performing a task, the digital assistant instantiates a process associated with the user device to perform the task. Instantiating a process includes invoking the process if the process is not already running. If at least one instance of the process is running, instantiating a process includes executing an existing instance of the process or generating a new instance of the process. For example, instantiating an audio configuration process includes invoking the audio configuration process, using an existing audio configuration process, or generating a new instance of the audio configuration process. In some examples, instantiating a process includes performing the task using the process. For example, as illustrated in FIG. 19A, in accordance with the user intent to increase the volume to its maximum level, the digital assistant instantiates an audio configuration process to set the volume to its maximum level. In some examples, the digital assistant further provides a spoken output 1972 such as "OK, I turned the volume all the way up."

[0343] With reference to FIG. 19B, the user device displays a user interface 1934 associated with performing a task. For example, the task includes viewing or editing a document. The user may desire to lower the screen brightness for eye protection. In some examples, the user provides a speech input 1954 such as "Set my screen brightness to 10% lower." The digital

assistant determines the user intent based on speech input 1954 and context information. For example, context information indicates that the current brightness configuration is at 90%. As a result, the digital assistant determines that the user intent is to reduce the brightness level from 90% to 80%. The digital assistant further determines whether the user intent indicates an informational request or a request for performing a task. For example, based on the determination that the user intent is to change the screen brightness to 80%, the digital assistant determines that a display configuration is to be varied, and therefore the user intent indicates a request for performing a task.

[0344] In some embodiments, in accordance with a determination that the user intent indicates a request for performing a task, the digital assistant instantiates a process to perform the task. For example, as illustrated in FIG. 19B, in accordance with the user intent to change the brightness level, the digital assistant instantiates a display configuration process to reduce the brightness level to 80%. In some examples, the digital assistant further provides a spoken output 1974 such as "OK, I turned your screen brightness to 80%." In some examples, as illustrated in FIG. 19B, the digital assistant provides an affordance 1936 enabling the user to manipulate a result of performing the task. For example, affordance 1936 can be a sliding bar allowing the user to further change the brightness level.

[0345] With reference to FIG. 19C, the user device displays a user interface 1938 associated with performing a task. For example, the task includes providing one or more notifications. A notification can include an alert of an email, a message, a reminder, or the like. In some examples, notifications are provided in user interface 1938. A notification can be displayed or provided to the user in real time or shortly after it is available at the user device. For example, a notification appears on user interface 1938 and/or user interface 1910 shorted after the user device receives it. Sometimes, the user may be performing an important task (e.g., editing a document) and may not want to be disturbed by the notifications. In some examples, the user provides a speech input 1956 such as "Don't notify me about incoming emails." The digital assistant determines that the user intent is to turn off the alert of emails. Based on the determination that the user intent is to turn off the alert of incoming emails, the digital assistant determines that a notification configuration is to be varied, and therefore the user intent indicates a request for performing a task.

[0346] In some embodiments, in accordance with a determination that the user intent indicates a request for performing a task, the digital assistant instantiates a process to perform the task. For example, as illustrated in FIG. 19C, in accordance with the user intent, the digital assistant instantiates a notification configuration process to turn off the alert of emails. In some examples, the digital assistant further provides a spoken output 1976 such as "OK, I turned off notifications for mail." In some examples, as illustrated in FIG. 19C, the digital assistant provides a user interface 1942 (e.g., a snippet or a window) enabling the user to manipulate a result of performing the task. For example, user interface 1942 provides an affordance 1943 (e.g., a cancel button). If the user desires to continue receiving notification of emails, for example, the user can select affordance 1943 to turn the notifications of emails back on. In some examples, the user can also provide another speech input, such as "Notify me of incoming emails" to turn on the notification of emails.

[0347] With reference to FIG. 19D, in some embodiments, the digital assistant may not be able to complete a task based on user's speech input and can thus provide a user interface to enable the user to perform the task. As shown in FIG. 19D, in some examples, the user provides a speech input 1958 such as "Show a custom message on my screen saver." The digital assistant determines that the user intent is to change the screen saver settings to show a custom message. The digital assistant further determines that the user intent is to vary a display configuration, and therefore the user intent indicates a request for performing a task.

[0348] In some embodiments, in accordance with a determination that the user intent indicates a request for performing a task, the digital assistant instantiates a process associated with the user device to perform the task. In some examples, if the digital assistant cannot complete the task based on the user intent, it provides a user interface enabling the user to perform the task. For example, based on speech input 1958, the digital assistant may not be able to determine the content of the custom message that is to be shown on the screen saver and therefore cannot complete the task of displaying the custom message. As illustrated in FIG. 19D, in some examples, the digital assistant instantiates a display configuration process and provides a user interface 1946 (e.g., a snippet or a window) to enable the user to manually change the screen saver settings. As another example, the digital assistant provides a link 1944 (e.g., a link to the display configurations) enabling the user to perform the task. The user selects link 1944 by

using an input device such as a mouse, a finger, or a stylus. Upon receiving the user's selection, the digital assistant instantiates a display configuration process and displays user interface 1946 to enable the user to change the screen saver settings. In some examples, the digital assistant further provides a spoken output 1978 such as "You can explore screen saver options in the screen saver configurations."

7. Process for Operating a Digital Assistant – Intelligent Search and Object Management.

[0349] FIGs. 20A-20G illustrate a flow diagram of an exemplary process 2000 for operating a digital assistant in accordance with some embodiments. Process 2000 may be performed using one or more devices 104, 108, 200, 400, or 600 (FIGs. 1, 2A, 4, or 6A-B). Operations in process 2000 are, optionally, combined or split, and/or the order of some operations is, optionally, changed.

[0350] With reference to FIG. 20A, at block 2002, prior to receiving a first speech input, an affordance to invoke a digital assistant service is displayed on a display associated with a user device. At block 2003, the digital assistant is invoked in response to receiving a pre-determined phrase. At block 2004, the digital assistant is invoked in response to receiving a selection of the affordance.

[0351] At block 2006, a first speech input is received from a user. At block 2008, context information associated with the user device is identified. At block 2009, the context information includes at least one of: user-specific data, metadata associated with one or more objects, sensor data, and user device configuration data.

[0352] At block 2010, a user intent is determined based on the first speech input and the context information. At block 2012, to determine the user intent, one or more actionable intents are determined. At block 2013, one or more parameters associated with the actionable intent are determined.

[0353] With reference to FIG. 20B, at block 2015, it is determined whether the user intent is to perform a task using a searching process or an object managing process. The searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device. At

block 2016, it is determined whether the speech input includes one or more keywords representing the searching process or the object managing process. At block 2018, it is determined whether the task is associated with searching. At block 2020, in accordance with a determination that the task is associated with searching, it is determined whether performing the task requires the searching process. At block 2021, in accordance with a determination that performing the task does not require the searching process, a spoken request to select the searching process or the object managing process is outputted, and a second speech input is received from the user. The second speech input indicates the selection of the searching process or the object managing process.

[0354] At block 2022, in accordance with a determination that performing the task does not require the searching process, it is determined, based on a pre-determined configuration, whether the task is to be performed using the searching process or the object managing process.

[0355] With reference to FIG. 20C, at block 2024, in accordance with a determination that the task is not associated with searching, it is determined whether the task is associated with managing at least one object. At block 2025, in accordance with a determination that the task is not associated with managing the at least one object, at least one of the following is performed: determining whether that task can be performed using a fourth process available to the user device and initiating a dialog with the user.

[0356] At block 2026, in accordance with a determination the user intent is to perform the task using the searching process, the task is performed using the searching process. At block 2028, at least one object is searched using the searching process. At block 2029, the at least one object includes at least one of a folder or a file. At block 2030, the file includes at least one of a photo, audio, or a video. At block 2031, the file is stored internally or externally to the user device. At block 2032, searching at least one of the folder or the file is based on metadata associated with the folder or the file. At block 2034, the at least one object includes a communication. At block 2035, the communication includes at least one of an email, a message, a notification, or a voicemail. At block 2036, metadata associated with the communication is searched.

[0357] With reference to FIG. 20D, at block 2037, the at least one object includes at least one of a contact or a calendar. At block 2038, the at least one object includes an application. At block 2039, the at least one object includes an online informational source.

[0358] At block 2040, in accordance with the determination that the user intent is to perform the task using the object managing process, the task is performed using the object managing process. At block 2042, the task is associated with searching, and the at least one object is searched using the object managing process. At block 2043, the at least one object includes at least one of a folder or a file. At block 2044, the file includes at least one of a photo, an audio, or a video. At block 2045, the file is stored internally or externally to the user device. At block 2046, searching at least one of the folder or the file is based on metadata associated with the folder or the file.

[0359] At block 2048, the object managing process is instantiated. Instantiating the object managing process includes invoking the object managing process, generating a new instance of the object managing process, or executing an existing instance of the object managing process.

[0360] With reference to FIG. 20E, at block 2049, the at least one object is created. At block 2050, the at least one object is stored. At block 2051, the at least one object is compressed. At block 2052, the at least one object is moved from a first physical or virtual storage to a second physical or virtual storage. At block 2053, the at least one object is copied from a first physical or virtual storage to a second physical or virtual storage. At block 2054, the at least one object stored in a physical or virtual storage is deleted. At block 2055, the at least one object stored at a physical or virtual storage is recovered. At block 2056, the at least one object is marked. Marking of the at least one object is at least one of visible or associated with metadata of the at least one object. At block 2057, the at least one object is backup according to a predetermined time period for backing up. At block 2058, the at least one object is shared among one or more electronic devices communicatively connected to the user device.

[0361] With reference to FIG. 20F, at block 2060, a response is provided based on a result of performing the task using the searching process or the object managing process. At block 2061, a first user interface is displayed providing the result of performing the task using the searching process or the object managing process. At block 2062, a link associated with the result of

performing the task using the searching process is displayed. At block 2063, a spoken output is provided according to the result of performing the task using the searching process or the object managing process.

[0362] At block 2064, it is provided an affordance that enables the user to manipulate the result of performing the task using the searching process or the object managing process. At block 2065, it is instantiated a third process that operates using the result of performing the task.

[0363] With reference to FIG. 20F, at block 2066, a confidence level is determined. At block 2067, the confidence level represents the accuracy in determining the user intent based on the first speech input and context information associated with the user device. At block 2068, the confidence level represents the accuracy in determining whether the user intent is to perform the task using the searching process or the object managing process.

[0364] With reference to FIG. 20G, at block 2069, the confidence level represents the accuracy in performing the task using the searching process or the object managing process.

[0365] At block 2070, the response is provided in accordance with the determination of the confidence level. At block 2071, it is determined whether the confidence level is greater than or equal to a threshold confidence level. At block 2072, in accordance with a determination that the confidence level is greater than or equal to the threshold confidence level, a first response is provided. At block 2073, in accordance with a determination that the confidence level is less than a threshold confidence level, a second response is provided.

8. Process for Operating a Digital Assistant – Continuity.

[0366] FIGs. 21A-21E illustrate a flow diagram of an exemplary process 2100 for operating a digital assistant in accordance with some embodiments. Process 2100 may be performed using one or more devices 104, 108, 200, 400, 600, 1400, 1500, 1600, or 1700 (FIGs. 1, 2A, 4, 6A-6B, 14A-14D, 15A-15D, 16A-16C, and 17A-17E). Operations in process 2100 are, optionally, combined or split and/or the order of some operations is, optionally, changed.

[0367] With reference to FIG. 21A, at block 2102, prior to receiving a first speech input, an affordance to invoke a digital assistant service is displayed on a display associated with a user

device. At block 2103, the digital assistant is invoked in response to receiving a pre-determined phrase. At block 2104, the digital assistant is invoked in response to receiving a selection of the affordance.

[0368] At block 2106, a first speech input is received from a user to perform a task. At block 2108, context information associated with the user device is identified. At block 2109, the user device is configured to provide a plurality of user interfaces. At block 2110, the user device includes a laptop computer, a desktop computer, or a server. At block 2112, the context information includes at least one of: user-specific data, metadata associated with one or more objects, sensor data, and user device configuration data.

[0369] At block 2114, a user intent is determined based on the speech input and the context information. At block 2115, to determine the user intent, one or more actionable intents are determined. At block 2116, one or more parameters associated with the actionable intent are determined.

[0370] With reference to FIG. 21B, at block 2118, in accordance with user intent, it is determined whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device. At block 2120, the first electronic device includes a laptop computer, a desktop computer, a server, a smartphone, a tablet, a set-top box, or a watch. At block 2121, determining whether the task is to be performed at the user device or at the first electronic device is based on one or more keywords included in the speech input. At block 2122, it is determined whether performing the task at the user device satisfies performance criteria. At block 2123, the performance criteria are determined based on one or more user preferences. At block 2124, the performance criteria are determined based on the device configuration data. At block 2125, the performance criteria are dynamically updated. At block 2126, in accordance with a determination that performing the task at the user device satisfies the performance criteria, it is determined that the task is to be performed at the user device.

[0371] With reference to FIG. 21C, at block 2128, in accordance with a determination that performing the task at the user device does not satisfy the performance criteria, it is determined whether performing the task at the first electronic device satisfies the performance criteria. At block 2130, in accordance with a determination that performing the task at the first electronic

device satisfies the performance criteria, it is determined that the task is to be performed at the first electronic device. At block 2132, in accordance with a determination that performing the task at the first electronic device does not meet the performance criteria, it is determined whether performing the task at the second electronic device satisfies the performance criteria.

[0372] At block 2134, in accordance with a determination that the task is to be performed at the user device and content for performing the task is located remotely, the content for performing the task is received. At block 2135, at least a portion of the content is received from the first electronic device. At least a portion of the content is stored in the first electronic device. At block 2136, at least a portion of the content is received from a third electronic device.

[0373] With reference to FIG. 21D, at block 2138, in accordance with a determination that the task is to be performed at the first electronic device and the content for performing the task is located remotely to the first electronic device, the content for performing the task is provided to the first electronic device. At block 2139, at least a portion of the content is provided from the user device to the first electronic device. At least a portion of the content is stored at the user device. At block 2140, at least a portion of the content is caused to be provided from a fourth electronic device to the first electronic device. At least a portion of the content is stored at the fourth electronic device.

[0374] At block 2142, the task is to be performed at the user device. A first response is provided at the user device using the received content. At block 2144, the task is performed at the user device. At block 2145, performing the task at the user device is a continuation of a task partially performed remotely to the user device. At block 2146, a first user interface is displayed associated with the task to be performed at the user device. At block 2148, a link associated with the task is to be performed at the user device. At block 2150, a spoken output is provided according to the task to be performed at the user device.

[0375] With reference to FIG. 21E, at block 2152, the task is to be performed at the first electronic device, and a second response is provided at the user device. At block 2154, the task is to be performed at the first electronic device. At block 2156, the task to be performed at the first electronic device is a continuation of a task performed remotely to the first electronic device. At block 2158, a spoken output is provided according to the task to be performed at the

first electronic device. At block 2160, a spoken output is provided according to the task to be performed at the first electronic device.

9. Process for Operating a Digital Assistant – System Configuration Management.

[0376] FIGs. 22A-22D illustrate a flow diagram of an exemplary process 2200 for operating a digital assistant in accordance with some embodiments. Process 2200 may be performed using one or more devices 104, 108, 200, 400, 600, or 1800 (FIGs. 1, 2A, 4, 6A-6B, and 18C-18D). Operations in process 2200 are, optionally, combined or split, and/or the order of some operations is, optionally, changed.

[0377] With reference to FIG. 22A, at block 2202, prior to receiving a speech input, an affordance to invoke a digital assistant service is displayed on a display associated with a user device. At block 2203, the digital assistant is invoked in response to receiving a pre-determined phrase. At block 2204, the digital assistant is invoked in response to receiving a selection of the affordance.

[0378] At block 2206, a speech input is received from a user to manage one or more system configurations of the user device. The user device is configured to concurrently provide a plurality of user interfaces. At block 2207, the one or more system configurations of the user device comprise audio configurations. At block 2208, the one or more system configurations of the user device comprise date and time configurations. At block 2209, the one or more system configurations of the user device comprise dictation configurations. At block 2210, the one or more system configurations of the user device comprise display configurations. At block 2211, the one or more system configurations of the user device comprise input device configurations. At block 2212, the one or more system configurations of the user device comprise network configurations. At block 2213, the one or more system configurations of the user device comprise notification configurations.

[0379] With reference to FIG. 22B, at block 2214, the one or more system configurations of the user device comprise printer configurations. At block 2215, the one or more system configurations of the user device comprise security configurations. At block 2216, the one or more system configurations of the user device comprise backup configurations. At block 2217,

the one or more system configurations of the user device comprise application configurations. At block 2218, the one or more system configurations of the user device comprise user interface configurations.

[0380] At block 2220, context information associated with the user device is identified. At block 2223, the context information comprises at least one of: user-specific data, device configuration data, and sensor data. At block 2224, the user intent is determined based on the speech input and the context information. At block 2225, one or more actionable intents are determined. At block 2226, one or more parameters associated with the actionable intent are determined.

[0381] With reference to FIG. 22C, at block 2228, it is determined whether the user intent indicates an informational request or a request for performing a task. At block 2229, it is determined whether the user intent is to vary a system configuration.

[0382] At block 2230, in accordance with a determination that the user intent indicates an informational request, a spoken response is provided to the informational request. At block 2231, status of one or more system configurations is obtained according to the informational request. At block 2232, the spoken response is provided according to the status of one or more system configurations.

[0383] At block 2234, in addition to providing the spoken response to the informational request, a first user interface is displayed to provide information according to the status of the one or more system configurations. At block 2236, in addition to providing the spoken response to the informational request, a link associated with the informational request is provided.

[0384] At block 2238, in accordance with a determination that the user intent indicates a request for performing a task, a process associated with the user device is instantiated to perform the task. At block 2239, the task is performed using the process. At block 2240, a first spoken output is provided according to a result of performing the task.

[0385] With reference to FIG. 22D, at block 2242, a second user interface is provided to enable the user to manipulate a result of performing the task. At block 2244, the second user interface comprises a link associated with the result of performing the task.

[0386] At block 2246, a third user interface is provided to enable the user to perform the task. At block 2248, the third user interface includes a link enabling the user to perform the task. At block 2250, a second spoken output associated with the third user interface is provided.

10. Electronic Device - intelligent search and object management

[0387] FIG. 23 shows a functional block diagram of electronic device 2300 configured in accordance with the principles of the various described examples, including those described with reference to FIGs. 8A-8F, 9A-9H, 10A-10B, 11A-11F, 12A-12D, 13A-13C, 14A-14D, 15A-15D, 16A-16C, 17A-17E, 18A-18F, and 19A-19D. The functional blocks of the device can be optionally implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described examples. It is understood by persons of skill in the art that the functional blocks described in FIG. 23 can be optionally combined or separated into sub-blocks to implement the principles of the various described examples. Therefore, the description herein optionally supports any possible combination, separation, or further definition of the functional blocks described herein.

[0388] As shown in FIG. 23, electronic device 2300 can include a microphone 2302 and processing unit 2308. In some examples, processing unit 2308 includes a receiving unit 2310, a an identifying unit 2312, a determining unit 2314, a performing unit 2316, a providing unit 2318, an instantiating unit 2320, a displaying unit 2322, an outputting unit 2324, an initiating unit 2326, a searching unit 2328, a generating unit 2330, an executing unit 2332, a creating unit 2334, an instantiating unit 2335, a storing unit 2336, a compressing unit 2338, a copying unit 2340, a deleting unit 2342, a recovering unit 2344, a marking unit 2346, a backing up unit 2348, a sharing unit 2350, a causing unit 2352, and an obtaining unit 2354.

[0389] In some examples, the processing unit 2308 is configured to receive (e.g., with the receiving unit 2310) a first speech input from a user; identify (e.g., with the identifying unit 2312) context information associated with the user device; and determine (e.g., with the determining unit 2314) a user intent based on the first speech input and the context information.

[0390] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether the user intent is to perform a task using a searching process or

an object managing process. The searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device.

[0391] In some examples, in accordance with a determination the user intent is to perform the task using the searching process, the processing unit 2308 is configured to perform (e.g., with the performing unit 2316) the task using the searching process. In some examples, in accordance with the determination that the user intent is to perform the task using the object managing process, the processing unit 2308 is configured to perform (e.g., with the performing unit 2316) the task using the object managing process.

[0392] In some examples, prior to receiving the first speech input, the processing unit 2308 is configured to display (e.g., with the displaying unit 2322), on a display associated with the user device, an affordance to invoke the digital assistant service.

[0393] In some examples, the processing unit 2308 is configured to invoke (e.g., with the invoking unit 2320) the digital assistant in response to receiving a pre-determined phrase.

[0394] In some examples, the processing unit 2308 is configured to invoke (e.g., with the invoking unit 2320) the digital assistant in response to receiving a selection of the affordance.

[0395] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) one or more actionable intents; and determine (e.g., with determining unit 2314) one or more parameters associated with the actionable intent.

[0396] In some examples, the context information comprises at least one of: user-specific data, metadata associated with one or more objects, sensor data, and user device configuration data.

[0397] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether the speech input includes one or more keywords representing the searching process or the object managing process.

[0398] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether the task is associated with searching. In accordance with a

determination that the task is associated with searching, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether performing the task requires the searching process; and in accordance with a determination that the task is not associated with searching, determine (e.g., with the determining unit 2314) whether the task is associated with managing at least one object.

[0399] In some examples, the task is associated with searching, and in accordance with a determination that performing the task does not require the searching process, the processing unit 2308 is configured to output (e.g., with the outputting unit 2324) a spoken request to select the searching process or the object managing process and receive (e.g., with the receiving unit 2310), from the user, a second speech input indicating the selection of the searching process or the object managing process.

[0400] In some examples, the task is associated with searching, and in accordance with a determination that performing the task does not require the searching process, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314), based on a pre-determined configuration, whether the task is to be performed using the searching process or the object managing process.

[0401] In some examples, the task is not associated with searching, and in accordance with a determination that the task is not associated with managing the at least one object, the processing unit 2308 is configured to perform (e.g., with the performing unit 2316) at least one of: determining (e.g., with the determining unit 2314) whether that task can be performed using a fourth process available to the user device; and initiating (e.g., with the initiating unit 2326) dialog with the user.

[0402] In some examples, the processing unit 2308 is configured to search (e.g., with the searching unit 2328) at least one object using the searching process.

[0403] In some examples, the at least one object includes at least one of a folder or a file. The file includes at least one of a photo, audio, or a video. The file is stored internally or externally to the user device.

[0404] In some examples, searching at least one of the folder or the file is based on metadata associated with the folder or the file.

[0405] In some examples, the at least one object includes a communication. The communication includes at least one of an email, a message, a notification, or a voicemail.

[0406] In some examples, the processing unit 2308 is configured to search (e.g., with the searching unit 2328) metadata associated with the communication.

[0407] In some examples, the at least one object includes at least one of a contact or a calendar.

[0408] In some examples, the at least one object includes an application.

[0409] In some examples, the at least one object includes an online informational source.

[0410] In some examples, the task is associated with searching, and the processing unit 2308 is configured to search (e.g., with the searching unit 2328) the at least one object using the object managing process.

[0411] In some examples, the at least one object includes at least one of a folder or a file. The file includes at least one of a photo, an audio, or a video. The file is stored internally or externally to the user device.

[0412] In some examples, searching at least one of the folder or the file is based on metadata associated with the folder or the file.

[0413] In some examples, the processing unit 2308 is configured to instantiate (e.g., with the instantiating unit 2335) the object managing process. Instantiating of the object managing process includes invoking the object managing process, generating a new instance of the object managing process, or executing an existing instance of the object managing process.

[0414] In some examples, the processing unit 2308 is configured to create (e.g., with the creating unit 2334) the at least one object.

[0415] In some examples, the processing unit 2308 is configured to store (e.g., with the storing unit 2336) the at least one object.

[0416] In some examples, the processing unit 2308 is configured to compress (e.g., with the compressing unit 2338) the at least one object.

[0417] In some examples, the processing unit 2308 is configured to move (e.g., with the moving unit 2339) the at least one object from a first physical or virtual storage to a second physical or virtual storage.

[0418] In some examples, the processing unit 2308 is configured to copy (e.g., with the copying unit 2340) the at least one object from a first physical or virtual storage to a second physical or virtual storage.

[0419] In some examples, the processing unit 2308 is configured to delete (e.g., with the deleting unit 2342) the at least one object stored in a physical or virtual storage.

[0420] In some examples, the processing unit 2308 is configured to recover (e.g., with the recovering unit 2344) at least one object stored at a physical or virtual storage.

[0421] In some examples, the processing unit 2308 is configured to mark (e.g., with the marking unit 2346) the at least one object. Marking of the at least one object is at least one of visible or associated with metadata of the at least one object.

[0422] In some examples, the processing unit 2308 is configured to back up (e.g., with the backing up unit 2348) the at least one object according to a predetermined time period for backing up.

[0423] In some examples, the processing unit 2308 is configured to share (e.g., with the sharing unit 2350) the at least one object among one or more electronic devices communicatively connected to the user device.

[0424] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a response based on a result of performing the task using the searching process or the object managing process.

[0425] In some examples, the processing unit 2308 is configured to display (e.g., with the displaying unit 2322) a first user interface providing the result of performing the task using the searching process or the object managing process.

[0426] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a link associated with the result of performing the task using the searching process.

[0427] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a spoken output according to the result of performing the task using the searching process or the object managing process.

[0428] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) an affordance that enables the user to manipulate the result of performing the task using the searching process or the object managing process.

[0429] In some examples, the processing unit 2308 is configured to instantiate (e.g., with the instantiating unit 2335) a third process that operates using the result of performing the task.

[0430] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) a confidence level; and provide (e.g., with providing unit 2318) the response in accordance with the determination of the confidence level.

[0431] In some examples, the confidence level represents the accuracy in determining the user intent based on the first speech input and context information associated with the user device.

[0432] In some examples, the confidence level represents the accuracy in determining whether the user intent is to perform the task using the searching process or the object managing process.

[0433] In some examples, the confidence level represents the accuracy in performing the task using the searching process or the object managing process.

[0434] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether the confidence level is greater than or equal to a threshold confidence level. In accordance with a determination that the confidence level is greater than or equal to the threshold confidence level, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a first response; and in accordance with a determination that the confidence level is less than a threshold confidence level, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a second response.

11. Electronic Device - Continuity

[0435] In some examples, the processing unit 2308 is configured to receive (e.g., with the receiving unit 2310) a speech input from a user to perform a task; identify (e.g., with the identifying unit 2312) context information associated with the user device; and determine (e.g., with the determining unit 2314) a user intent based on the speech input and context information associated with the user device.

[0436] In some examples, the processing unit 2308 is configured to, in accordance with user intent, determine (e.g., with the determining unit 2314) whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device.

[0437] In some examples, in accordance with a determination that the task is to be performed at the user device and content for performing the task is located remotely, the processing unit 2308 is configured to receive (e.g., with the receiving unit 2310) the content for performing the task.

[0438] In some examples, in accordance with a determination that the task is to be performed at the first electronic device and the content for performing the task is located remotely to the first electronic device, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) the content for performing the task to the first electronic device.

[0439] In some examples, the user device is configured to provide a plurality of user interfaces.

[0440] In some examples, the user device includes a laptop computer, a desktop computer, or a server.

[0441] In some examples, the first electronic device includes a laptop computer, a desktop computer, a server, a smartphone, a tablet, a set-top box, or a watch.

[0442] In some examples, the processing unit 2308 is configured to, prior to receiving the speech input, display (e.g., with the displaying unit 2322), on a display of the user device, an affordance to invoke the digital assistant.

[0443] In some examples, the processing unit 2308 is configured to invoke (e.g., with the invoking unit 2320) the digital assistant in response to receiving a pre-determined phrase.

[0444] In some examples, the processing unit 2308 is configured to invoke (e.g., with the invoking unit 2320) the digital assistant in response to receiving a selection of the affordance.

[0445] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) one or more actionable intents; and determine (e.g., with the determining unit 2314) one or more parameters associated with the actionable intent.

[0446] In some examples, the context information comprises at least one of: user-specific data, sensor data, and user device configuration data.

[0447] In some examples, determining whether the task is to be performed at the user device or at the first electronic device is based on one or more keywords included in the speech input.

[0448] In some examples, the processing unit 2308 is configured to determine (e.g., with determining unit 2314) whether performing the task at the user device satisfies performance criteria.

[0449] In some examples, in accordance with a determination that performing the task at the user device satisfies the performance criteria, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) that the task is to be performed at the user device.

[0450] In some examples, in accordance with a determination that performing the task at the user device does not satisfy the performance criteria, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether performing the task at the first electronic device satisfies the performance criteria.

[0451] In some examples, in accordance with a determination that performing the task at the first electronic device satisfies the performance criteria, the processing unit 2308 is configured to determine (e.g., with the determining 2314) that the task is to be performed at the first electronic device.

[0452] In some examples, in accordance with a determination that the performing the task at the first electronic device does not meet the performance criteria, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether performing the task at the second electronic device satisfies the performance criteria.

[0453] In some examples, the performance criteria are determined based on one or more user preferences.

[0454] In some examples, the performance criteria are determined based on the device configuration data.

[0455] In some examples, the performance criteria are dynamically updated.

[0456] In some examples, in accordance with a determination that the task is to be performed at the user device and content for performing the task is located remotely, the processing unit 2308 is configured to receive (e.g., with the receiving unit 2310) at least a portion of the content from the first electronic device, wherein at least a portion of the content is stored in the first electronic device.

[0457] In some examples, in accordance with a determination that the task is to be performed at the user device and content for performing the task is located remotely, the processing unit 2308 is configured to receive (e.g., with the receiving unit 2310) at least a portion of the content from a third electronic device.

[0458] In some examples, in accordance with a determination that the task is to be performed at the first electronic device and the content for performing the task is located remotely to the first electronic device, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) at least a portion of the content from the user device to the first electronic device, wherein at least a portion of the content is stored at the user device.

[0459] In some examples, in accordance with a determination that the task is to be performed at the first electronic device and the content for performing the task is located remotely to the first electronic device, the processing unit 2308 is configured to cause (e.g., with the causing unit 2352) at least a portion of the content to be provided from a fourth electronic device to the first electronic device. At least a portion of the content is stored at the fourth electronic device.

[0460] In some examples, the task is to be performed at the user device, and processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a first response at the user device using the received content.

[0461] In some examples, the processing unit 2308 is configured to perform (e.g., with the performing unit 2316) the task at the user device.

[0462] In some examples, performing the task at the user device is a continuation of a task partially performed remotely to the user device.

[0463] In some examples, the processing unit 2308 is configured to display (e.g., with the displaying unit 2322) a first user interface associated with the task to be performed at the user device.

[0464] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a link associated with the task to be performed at the user device.

[0465] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a spoken output according to the task to be performed at the user device.

[0466] In some examples, the task is to be performed at the first electronic device, and the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a second response at the user device.

[0467] In some examples, the processing unit 2308 is configured to cause (e.g., with the causing unit 2352) the task to be performed at the first electronic device.

[0468] In some examples, the task to be performed at the first electronic device is a continuation of a task performed remotely to the first electronic device.

[0469] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a spoken output according to the task to be performed at the first electronic device.

[0470] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) an affordance that enables the user to select another electronic device for performance of the task.

12. Electronic Device – System Configuration Management

[0471] In some examples, the processing unit 2308 is configured to receive (e.g., with the receiving unit 2310) a speech input from a user to manage one or more system configurations of the user device. The user device is configured to concurrently provide a plurality of user interfaces.

[0472] In some examples, the processing unit 2308 is configured to identify (e.g., with the identifying unit 2312) context information associated with the user device; and determine (e.g., with the determining unit 2314) a user intent based on the speech input and context information.

[0473] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether the user intent indicates an informational request or a request for performing a task.

[0474] In some examples, in accordance with a determination that the user intent indicates an informational request, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a spoken response to the informational request.

[0475] In some examples, in accordance with a determination that the user intent indicates a request for performing a task, the processing unit 2308 is configured to instantiate (e.g., with the instantiating unit 2335) a process associated with the user device to perform the task.

[0476] In some examples, the processing unit 2308 is configured to, prior to receiving the speech input, display (e.g., with the displaying unit 2322) on a display of the user device, an affordance to invoke the digital assistant.

[0477] In some examples, the processing unit 2308 is configured to invoke (e.g., with the invoking unit 2320) the digital assistant service in response to receiving a pre-determined phrase.

[0478] In some examples, the processing unit 2308 is configured to invoke (e.g., with the invoking unit 2320) the digital assistant service in response to receiving a selection of the affordance.

[0479] In some examples, the one or more system configurations of the user device comprise audio configurations.

[0480] In some examples, the one or more system configurations of the user device comprise date and time configurations.

[0481] In some examples, the one or more system configurations of the user device comprise dictation configurations.

[0482] In some examples, the one or more system configurations of the user device comprise display configurations.

[0483] In some examples, the one or more system configurations of the user device comprise input device configurations.

[0484] In some examples, the one or more system configurations of the user device comprise network configurations.

[0485] In some examples, the one or more system configurations of the user device comprise notification configurations.

[0486] In some examples, the one or more system configurations of the user device comprise printer configurations.

[0487] In some examples, the one or more system configurations of the user device comprise security configurations.

[0488] In some examples, the one or more system configurations of the user device comprise backup configurations.

[0489] In some examples, the one or more system configurations of the user device comprise application configurations.

[0490] In some examples, the one or more system configurations of the user device comprise user interface configurations.

[0491] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) one or more actionable intents; and determine (e.g., with the determining unit 2314) one or more parameters associated with the actionable intent.

[0492] In some examples, the context information comprises at least one of: user-specific data, device configuration data, and sensor data.

[0493] In some examples, the processing unit 2308 is configured to determine (e.g., with the determining unit 2314) whether the user intent is to vary a system configuration.

[0494] In some examples, the processing unit 2308 is configured to obtain (e.g., with the obtaining unit 2354) status of one or more system configurations according to the informational request; and provide (e.g., with the providing unit 2318) the spoken response according to the status of one or more system configurations.

[0495] In some examples, in accordance with a determination that the user intent indicates an informational request, the processing unit 2308 is configured to, in addition to providing the spoken response to the informational request, display (e.g., with the displaying unit 2322) a first user interface providing information according to the status of the one or more system configurations.

[0496] In some examples, in accordance with a determination that the user intent indicates an informational request, the processing unit 2308 is configured to, in addition to providing the spoken response to the informational request, provide (e.g., with the providing unit 2318) a link associated with the informational request.

[0497] In some examples, in accordance with a determination that the user intent indicates a request for performing a task, the processing unit 2308 is configured to perform (e.g., with the performing unit 2316) the task using the process.

[0498] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a first spoken output according to a result of performing the task.

[0499] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a second user interface enabling the user to manipulate a result of performing the task.

[0500] In some examples, the second user interface comprises a link associated with the result of performing the task.

[0501] In some examples, in accordance with a determination that the user intent indicates a request for performing a task, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a third user interface enabling the user to perform the task.

[0502] In some examples, the third user interface includes a link enabling the user to perform the task.

[0503] In some examples, the processing unit 2308 is configured to provide (e.g., with the providing unit 2318) a second spoken output associated with the third user interface.

[0504] The operation described above with respect to FIG. 23 is, optionally, implemented by components depicted in FIGs. 1, 2A, 4, 6A-B, or 7A-7B. For example, receiving operation 2310, identifying operation 2312, determining operation 2314, performing operation 2316, and providing operation 2318 are optionally implemented by processor(s) 220. 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, 2A, 4, 6A-B, or 7A-7B.

[0505] It is understood by persons of skill in the art that the functional blocks described in FIG. 12 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, processing unit 2308 can have an associated "controller" unit that is operatively coupled with processing unit 2308 to enable operation. This controller unit is not separately illustrated in FIG. 23 but is understood to be within the grasp of one of ordinary skill in the art who is designing a device having a processing unit 2308, such as device 2300. As another example, one or more units, such as the receiving unit 2310, may be hardware units outside of processing unit 2308 in some embodiments. The description herein thus optionally supports combination, separation, and/or further definition of the functional blocks described herein.

[0506] The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the techniques and their practical applications. Others skilled in the art are thereby enabled to best utilize the techniques and various embodiments with various modifications as are suited to the particular use contemplated.

[0507] Although the disclosure and examples have been fully described with reference to the accompanying drawings, 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 claims.

WHAT IS CLAIMED IS:

1. An electronic device, comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for:

receiving a first speech input from a user;

identifying context information associated with the user device;

determining a user intent based on the first speech input and the context information;

determining whether the user intent is to perform a task using a searching process or an object managing process, wherein the searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device;

in accordance with a determination the user intent is to perform the task using the searching process, performing the task using the searching process; and

in accordance with the determination that the user intent is to perform the task using the object managing process, performing the task using the object managing process.

2. The electronic device of claim 1, further comprising, prior to receiving the first speech input:

invoking the digital assistant in response to receiving a pre-determined phrase, or

invoking the digital assistant in response to receiving a selection of an affordance.

3. The electronic device of claim 1, wherein determining the user intent comprises:

determining one or more actionable intents; and

determining one or more parameters associated with the actionable intent.

4. The electronic device of claim 1, wherein the context information comprises at least one of: user-specific data, metadata associated with one or more objects, sensor data, and user device configuration data.

5. The electronic device of claim 1, wherein determining whether the user intent is to perform the task using the searching process or the object managing process comprises:

determining whether the speech input includes one or more keywords representing the searching process or the object managing process.

6. The electronic device of claim 1, wherein determining whether the user intent is to perform the task using the searching process or the object managing process comprises:

determining whether the task is associated with searching;

in accordance with a determination that the task is associated with searching, determining whether performing the task requires the searching process; and

in accordance with a determination that the task is not associated with searching, determining whether the task is associated with managing at least one object.

7. The electronic device of claim 6, wherein the task is associated with searching, further comprising:

in accordance with a determination that performing the task does not require the searching process,

outputting a spoken request to select the searching process or the object managing process, and

receiving, from the user, a second speech input indicating the selection of the searching process or the object managing process.

8. The electronic device of claim 6, wherein the task is associated with searching, further comprising:

in accordance with a determination that performing the task does not require the searching process, determining, based on a pre-determined configuration, whether the task is to be performed using the searching process or the object managing process.

9. The electronic device of claim 6, wherein the task is not associated with searching, further comprising:

in accordance with a determination that the task is not associated with managing the at least one object, performing at least one of:

determining whether that task can be performed using a fourth process available to the user device; and

initiating a dialog with the user.

10. The electronic device of claim 1, wherein performing the task using the searching process comprises:

searching at least one object using the searching process.

11. The electronic device of claim 10, wherein the at least one object includes at least one of: a folder, a file, a communication, a contact, a calendar, an application, and an online informational source.

12. The electronic device of claim 11, wherein the file includes at least one of a photo, audio, or a video; wherein the file is stored internally or externally to the user device; and wherein searching at least one of the folder or the file is based on metadata associated with the folder or the file.

13. The electronic device of claim 11, wherein the communication includes at least one of an email, a message, a notification, or a voicemail.

14. The electronic device of claim 11, further comprising searching metadata associated with the communication.

15. The electronic device of claim 1, wherein the task is associated with searching, and wherein performing the task using the object managing process comprises:

searching the at least one object using the object managing process.

16. The electronic device of claim 15, wherein the at least one object includes at least one of a folder or a file.

17. The electronic device of claim 16, wherein the file includes at least one of a photo, an audio, or a video; wherein the file is stored internally or externally to the user device; and wherein searching at least one of the folder or the file is based on metadata associated with the folder or the file.

18. The electronic device of claim 1, wherein performing the task using the object managing process comprises instantiating the object managing process, wherein instantiating the object managing process comprises invoking the object managing process, generating a new instance of the object managing process, or executing an existing instance of the object managing process.

19. The electronic device of claim 1, wherein performing the task using the object managing process comprises at least one of:

creating the at least one object;

storing the at least one object;

compressing the at least one object;

moving the at least one object from a first physical or virtual storage to a second physical or virtual storage;

copying the at least one object from a first physical or virtual storage to a second physical or virtual storage;

deleting the at least one object stored in a physical or virtual storage;

recovering at least one object stored at a physical or virtual storage;

marking the at least one object, wherein marking of the at least one object is at least one of visible or associated with metadata of the at least one object;

backing up the at least one object according to a predetermined time period for backing up; and

sharing the at least one object among one or more electronic devices communicatively connected to the user device.

20. The electronic device of claim 1, further comprising: providing a response based on a result of performing the task using the searching process or the object managing process.

21. The electronic device of claim 20, wherein providing the response based on the result of performing the task using the searching process or the object managing process comprises at least one of:

displaying a first user interface providing the result of performing the task using the searching process or the object managing process;

providing a link associated with the result of performing the task using the searching process;

providing a spoken output according to the result of performing the task using the searching process or the object managing process;

providing an affordance that enables the user to manipulate the result of performing the task using the searching process or the object managing process; and

instantiating a third process that operates using the result of performing the task.

22. The electronic device of claim 20, wherein providing the response based on the result of performing the task using searching process or the object managing process comprises:

determining a confidence level; and

providing the response in accordance with the determination of the confidence level.

23. The electronic device of claim 22, wherein the confidence level represents at least one of:

the accuracy in determining the user intent based on the first speech input and context information associated with the user device;

the accuracy in determining whether the user intent is to perform the task using the searching process or the object managing process; and

the accuracy in performing the task using the searching process or the object managing process.

24. The electronic device of claim 22, wherein providing the response in accordance with the determination of the confidence level comprises:

determining whether the confidence level is greater than or equal to a threshold confidence level;

in accordance with a determination that the confidence level is greater than or equal to the threshold confidence level, providing a first response; and

in accordance with a determination that the confidence level is less than a threshold confidence level, providing a second response.

25. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors of an electronic device, cause the electronic device to:

receive a first speech input from a user;

identify context information associated with the user device;

determine a user intent based on the first speech input and the context information;

determine whether the user intent is to perform a task using a searching process or an object managing process, wherein the searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device;

in accordance with a determination the user intent is to perform the task using the searching process, perform the task using the searching process; and

in accordance with the determination that the user intent is to perform the task using the object managing process, perform the task using the object managing process.

26. A method for providing a digital assistant service, comprising:

receiving a first speech input from a user;

identifying context information associated with the user device;

determining a user intent based on the first speech input and the context information;

determining whether the user intent is to perform a task using a searching process or an object managing process, wherein the searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device;

in accordance with a determination the user intent is to perform the task using the searching process, performing the task using the searching process; and

in accordance with the determination that the user intent is to perform the task using the object managing process, performing the task using the object managing process.

ABSTRACT

Systems and processes for operating a digital assistant are provided. In one example, a method includes receiving a first speech input from a user. The method further includes identifying context information and determining a user intent based on the first speech input and the context information. The method further includes determining whether the user intent is to perform a task using a searching process or an object managing process. The searching process is configured to search data, and the object managing process is configured to manage objects. The method further includes, in accordance with a determination the user intent is to perform the task using the searching process, performing the task using the searching process; and in accordance with the determination that the user intent is to perform the task using the object managing process, performing the task using the object managing process.

DK 2017 70434 A1

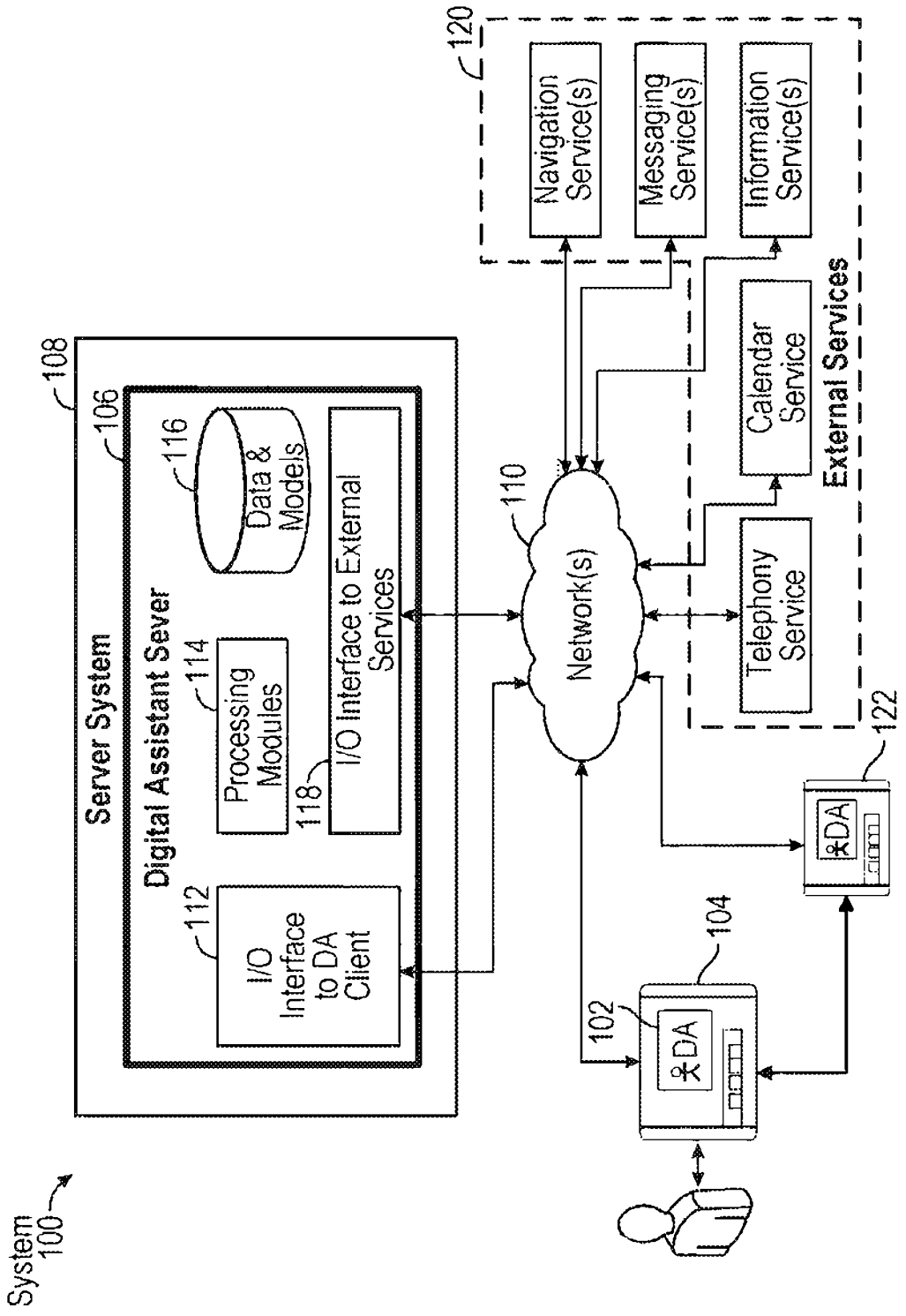


FIG. 1

DK 2017 70434 A1

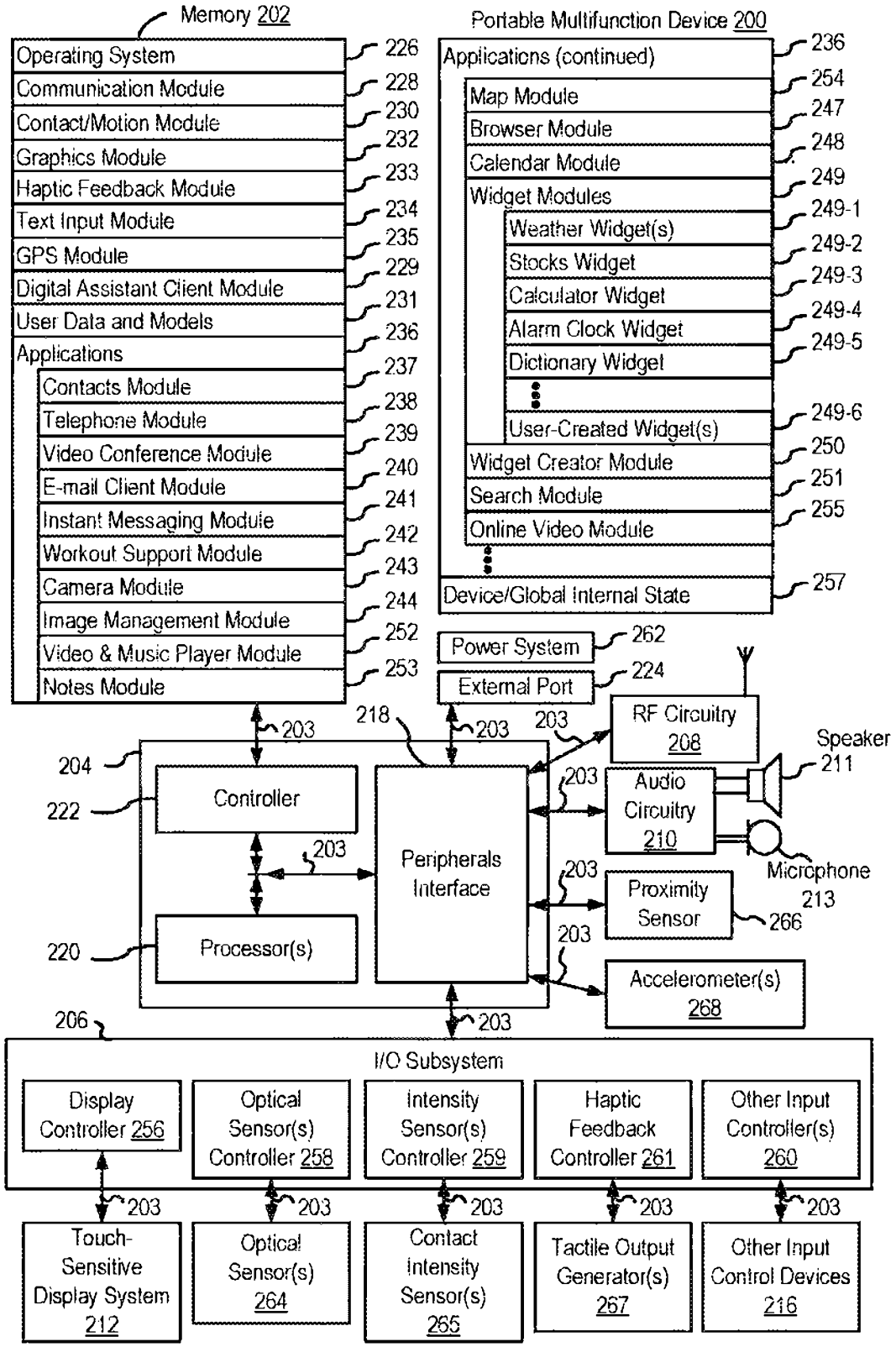


FIG. 2A

DK 2017 70434 A1

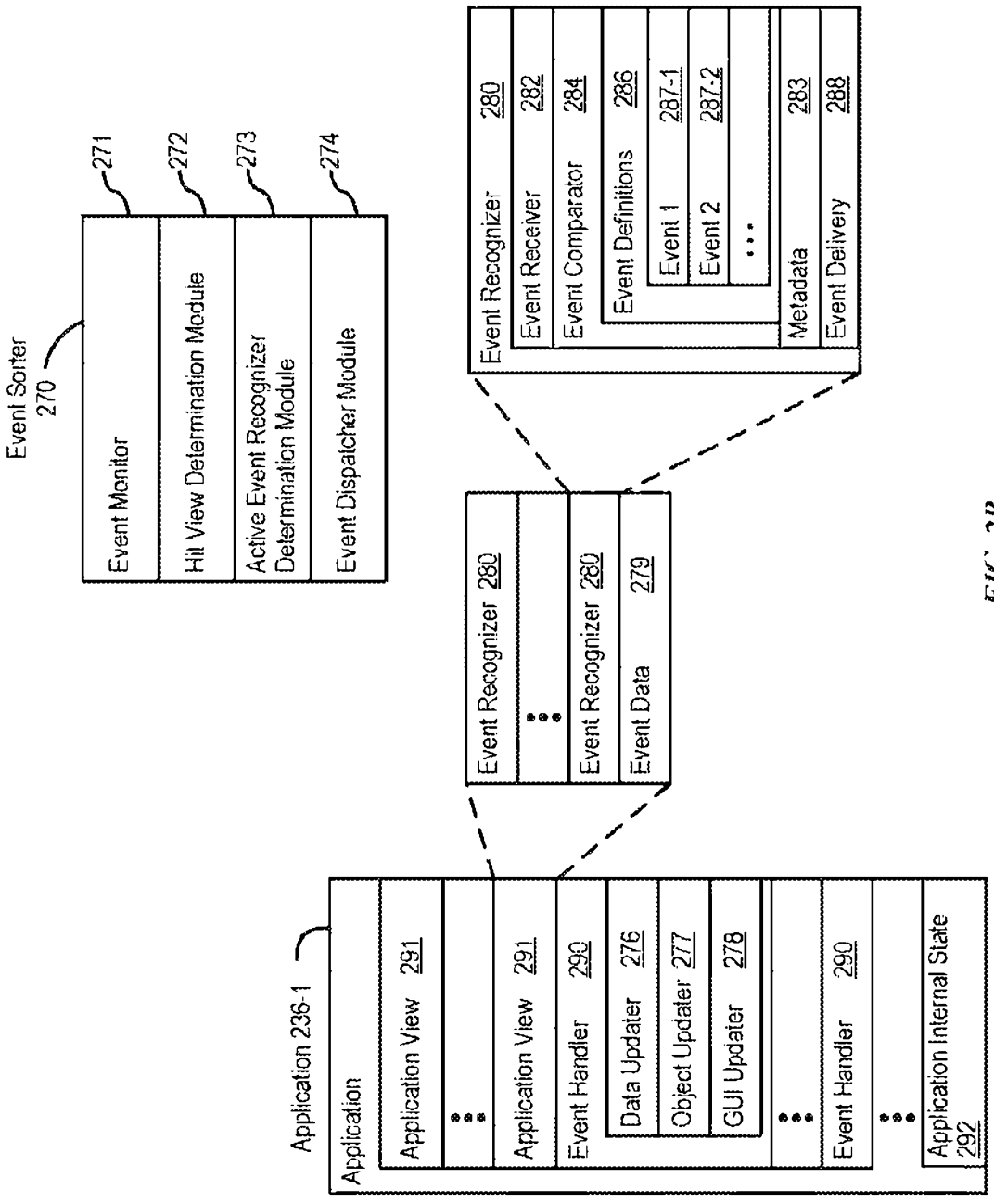


FIG. 2B

DK 2017 70434 A1

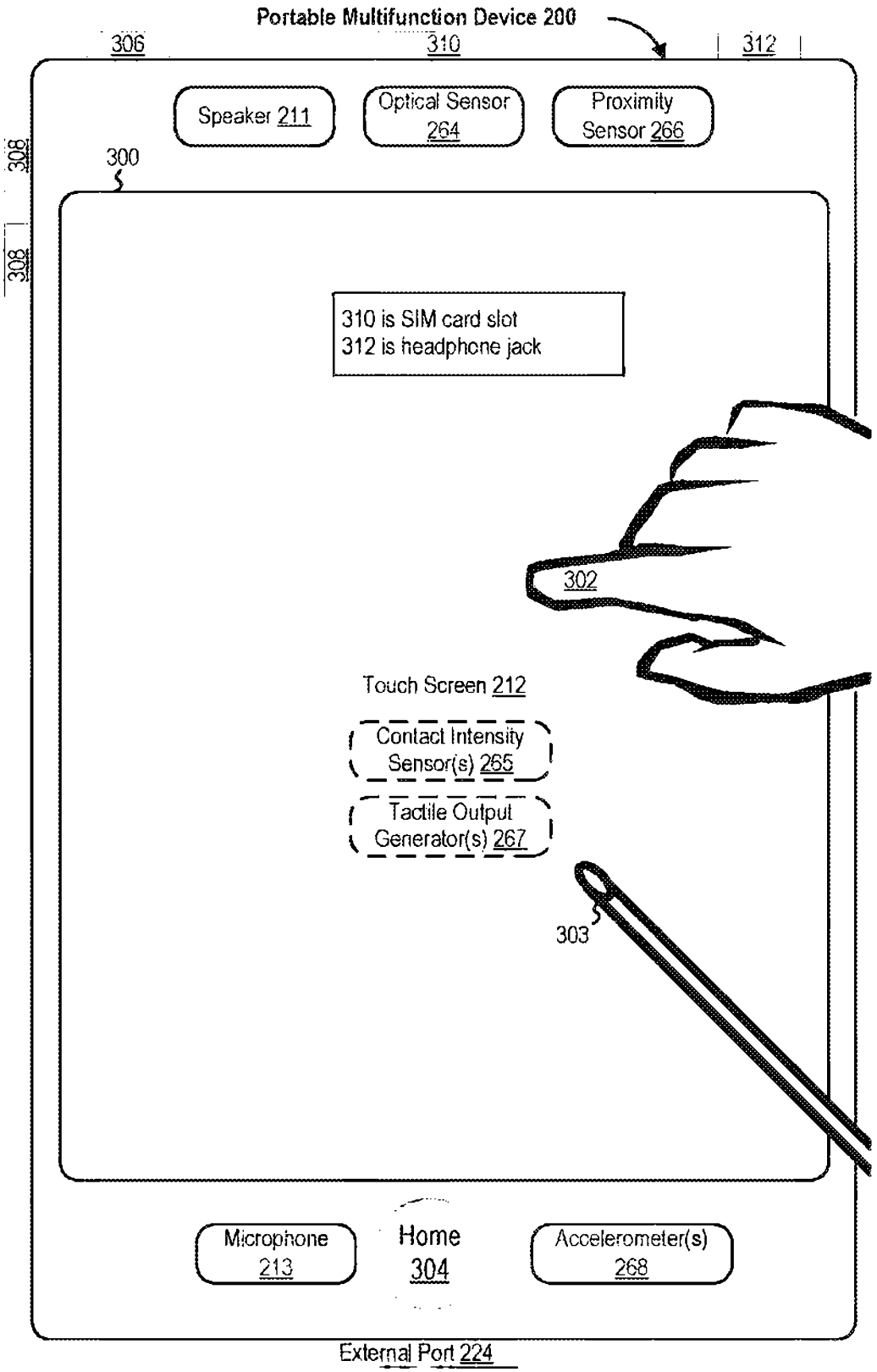


FIG. 3

DK 2017 70434 A1

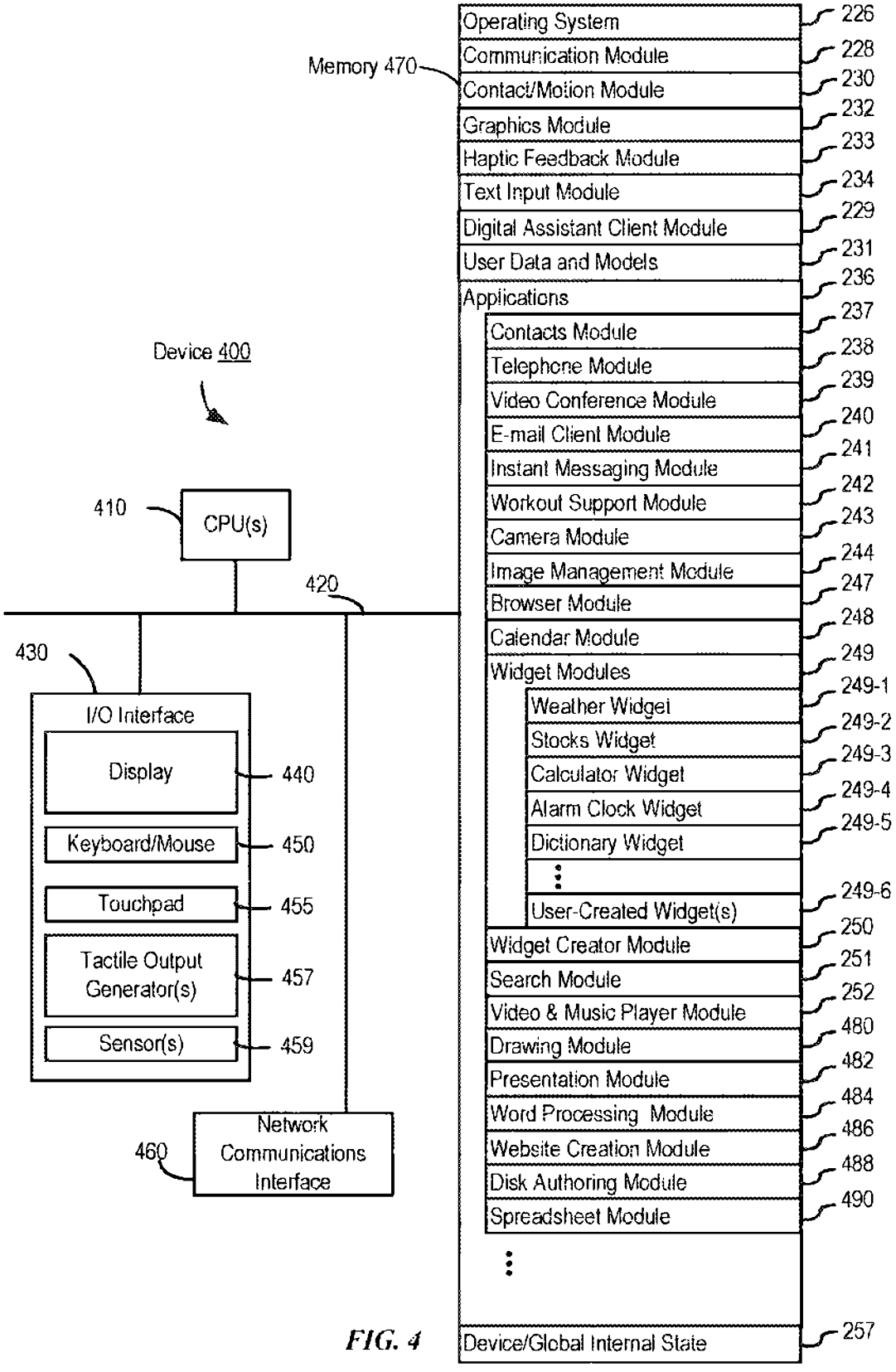


FIG. 4

DK 2017 70434 A1

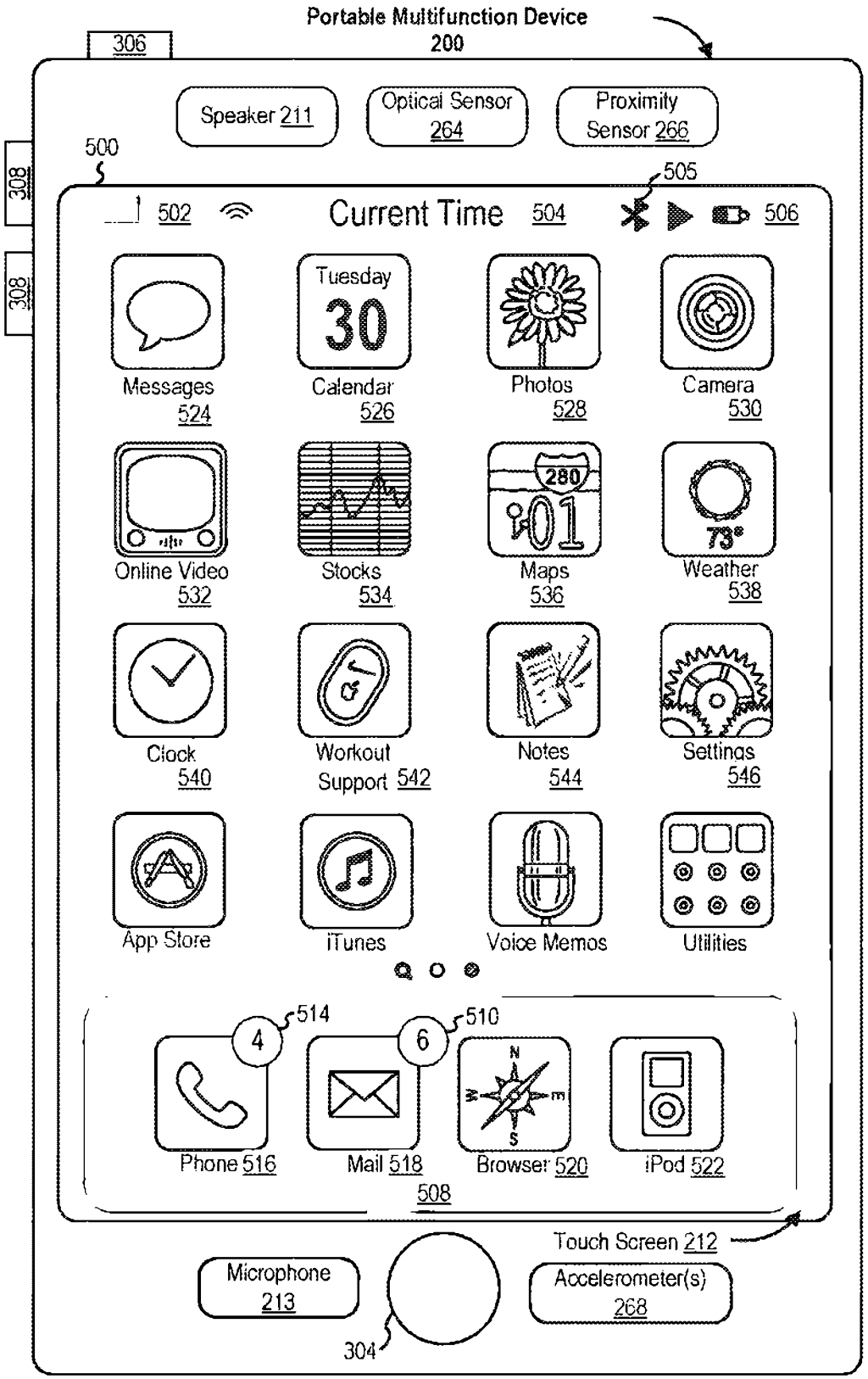
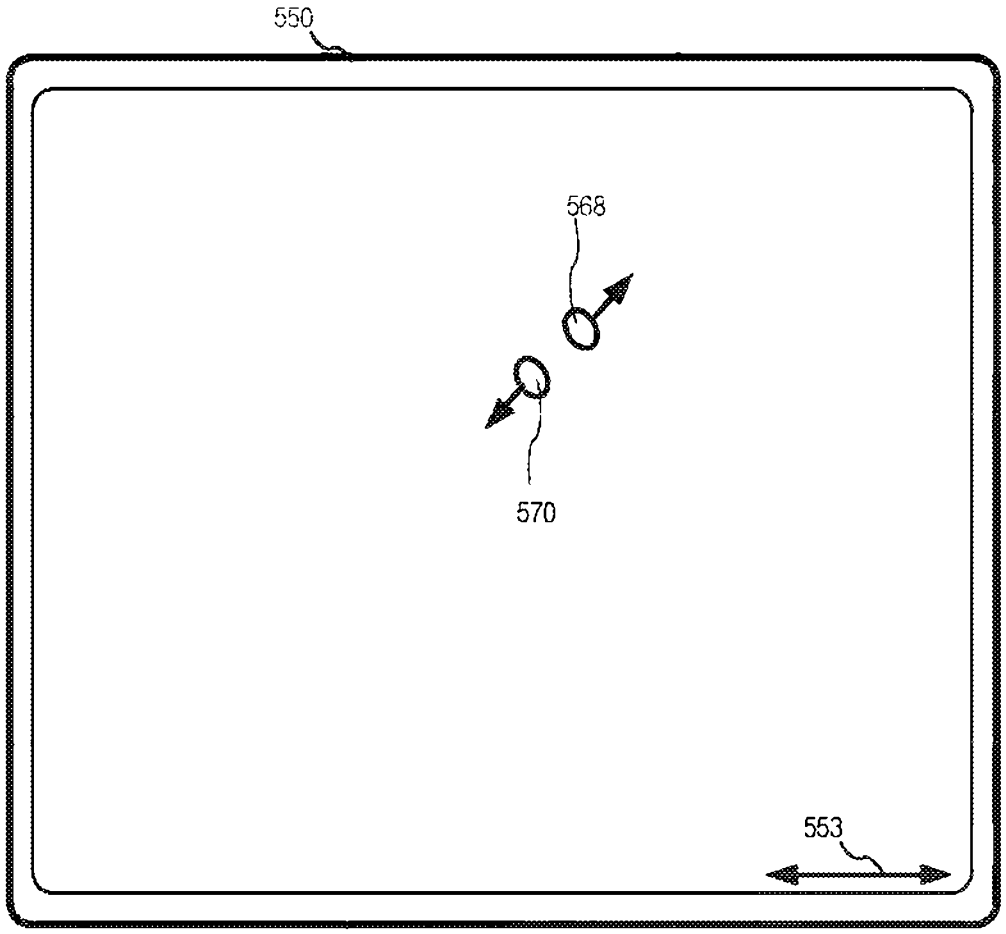


FIG. 5A

DK 2017 70434 A1



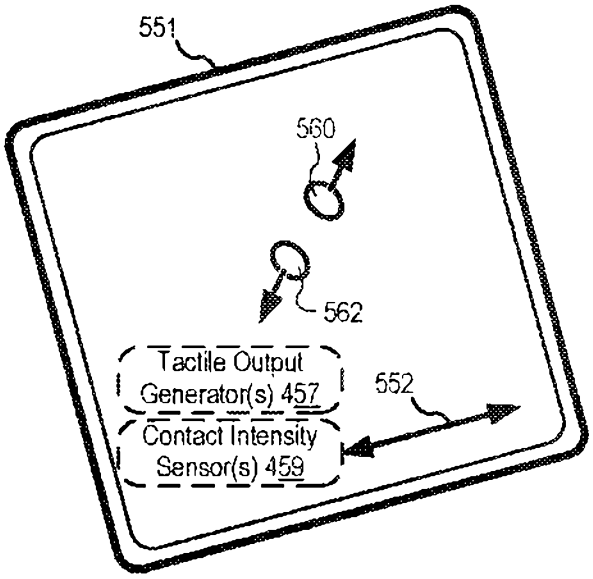


FIG. 5B

DK 2017 70434 A1

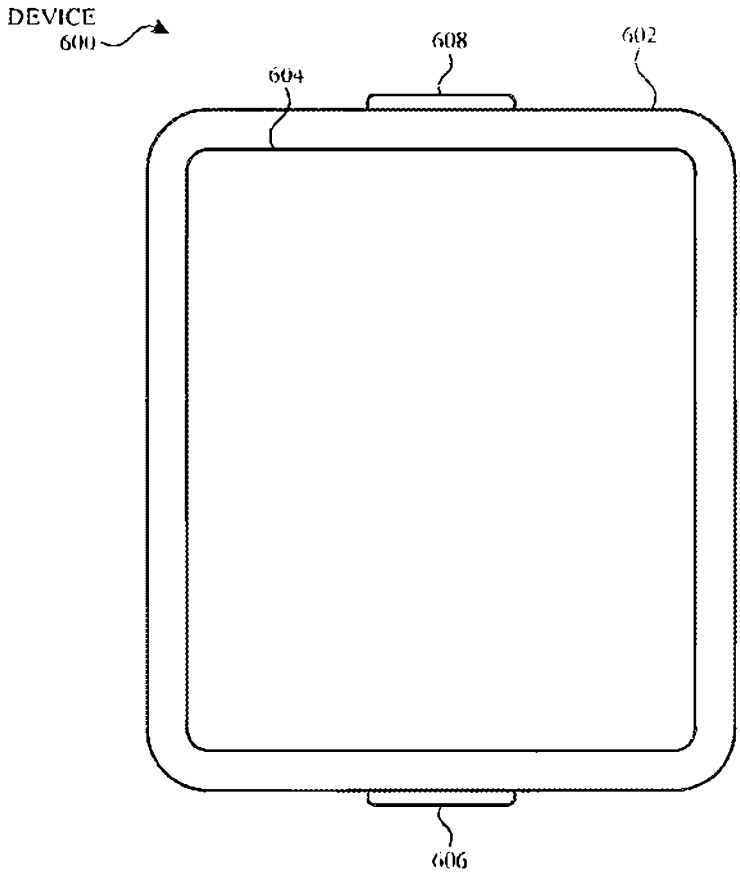


FIG. 6A

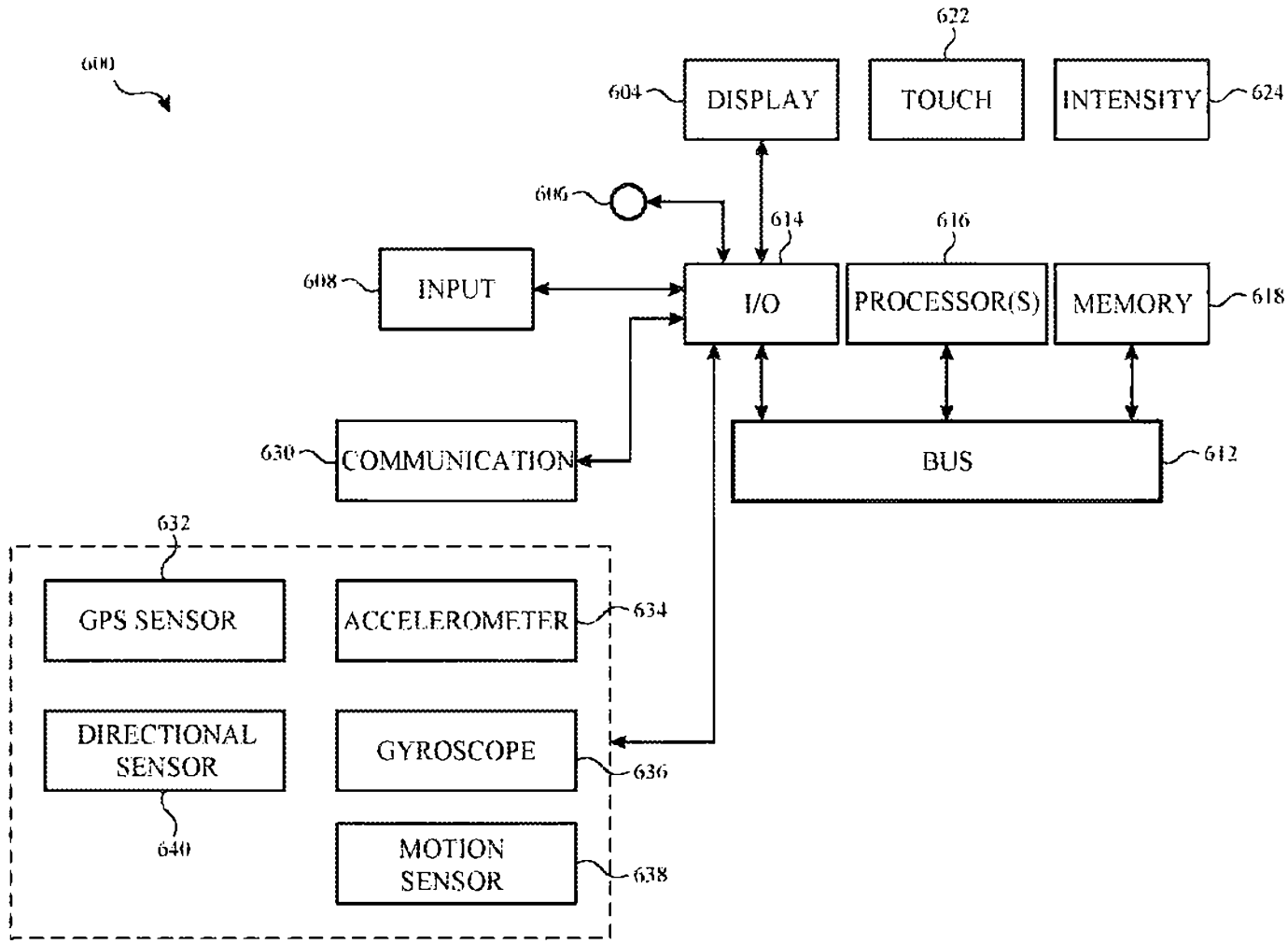

600
622
604
DISPLAY
TOUCH
INTENSITY
624
606
614
616
608
INPUT
I/O
PROCESSOR(S)
MEMORY
618
630
COMMUNICATION
BUS
612
632
GPS SENSOR
ACCELEROMETER
634
DIRECTIONAL SENSOR
GYROSCOPE
636
640
MOTION SENSOR
638

FIG. 6B

DK 2017 70434 A1

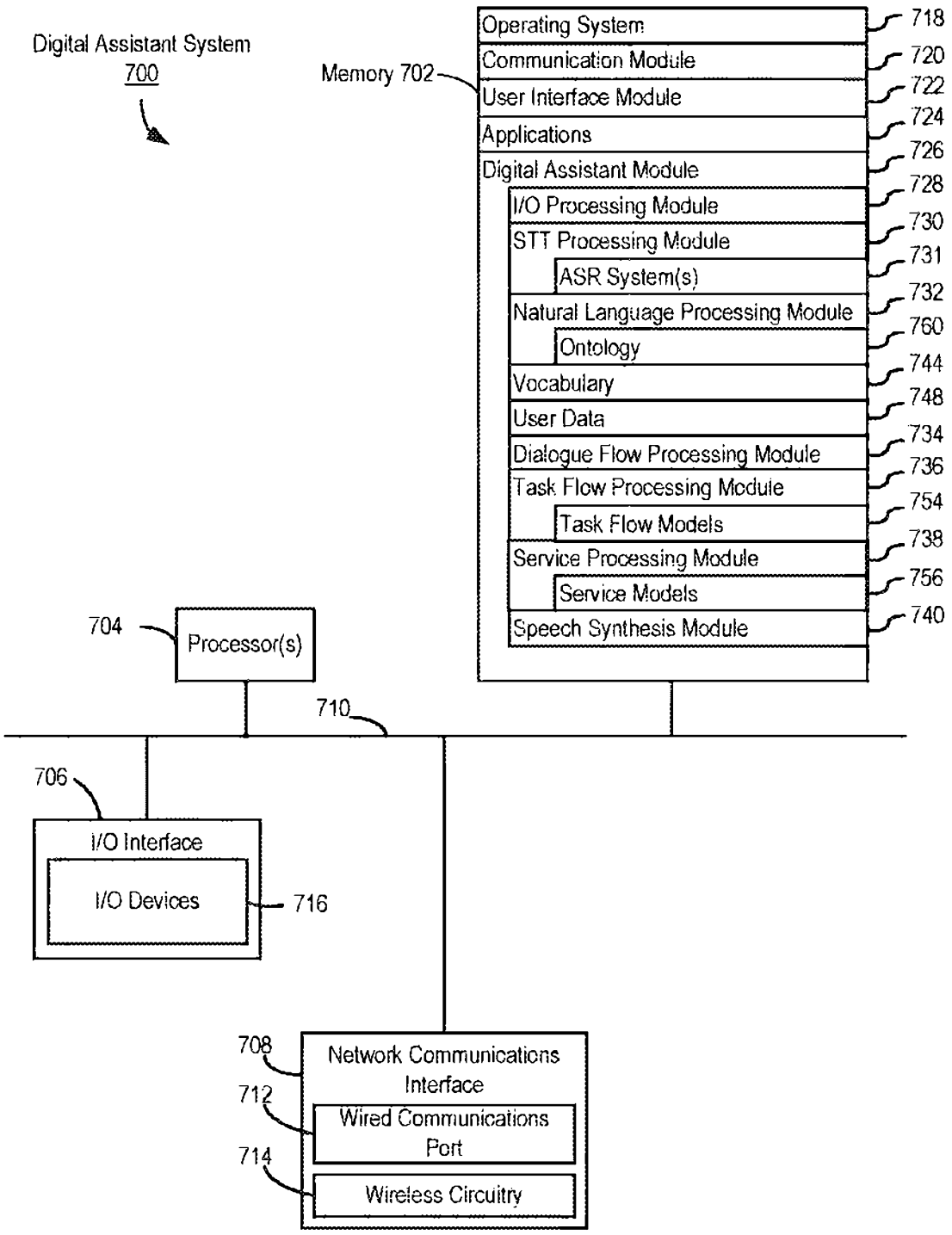


FIG. 7A

DK 2017 70434 A1

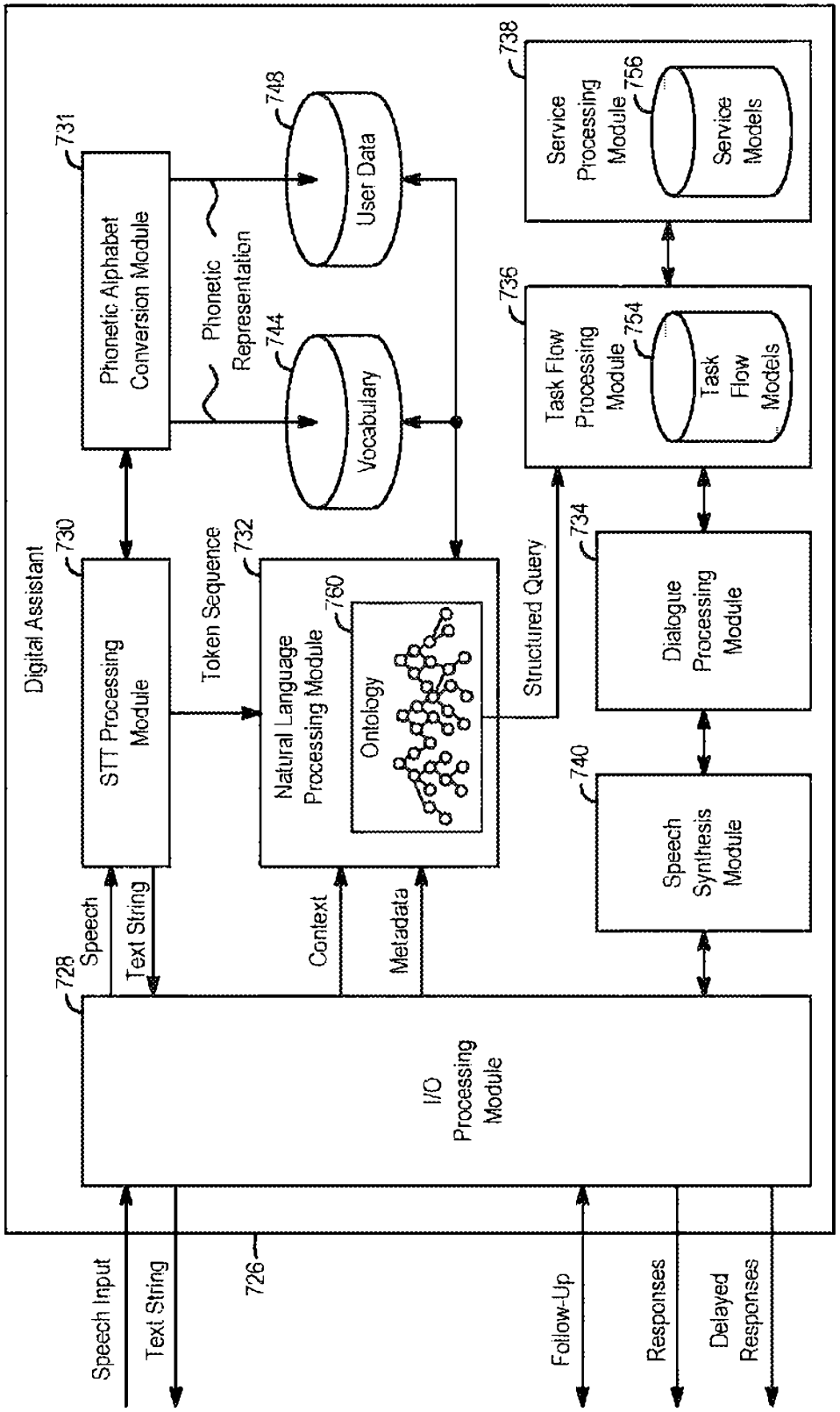


FIG. 7B

DK 2017 70434 A1

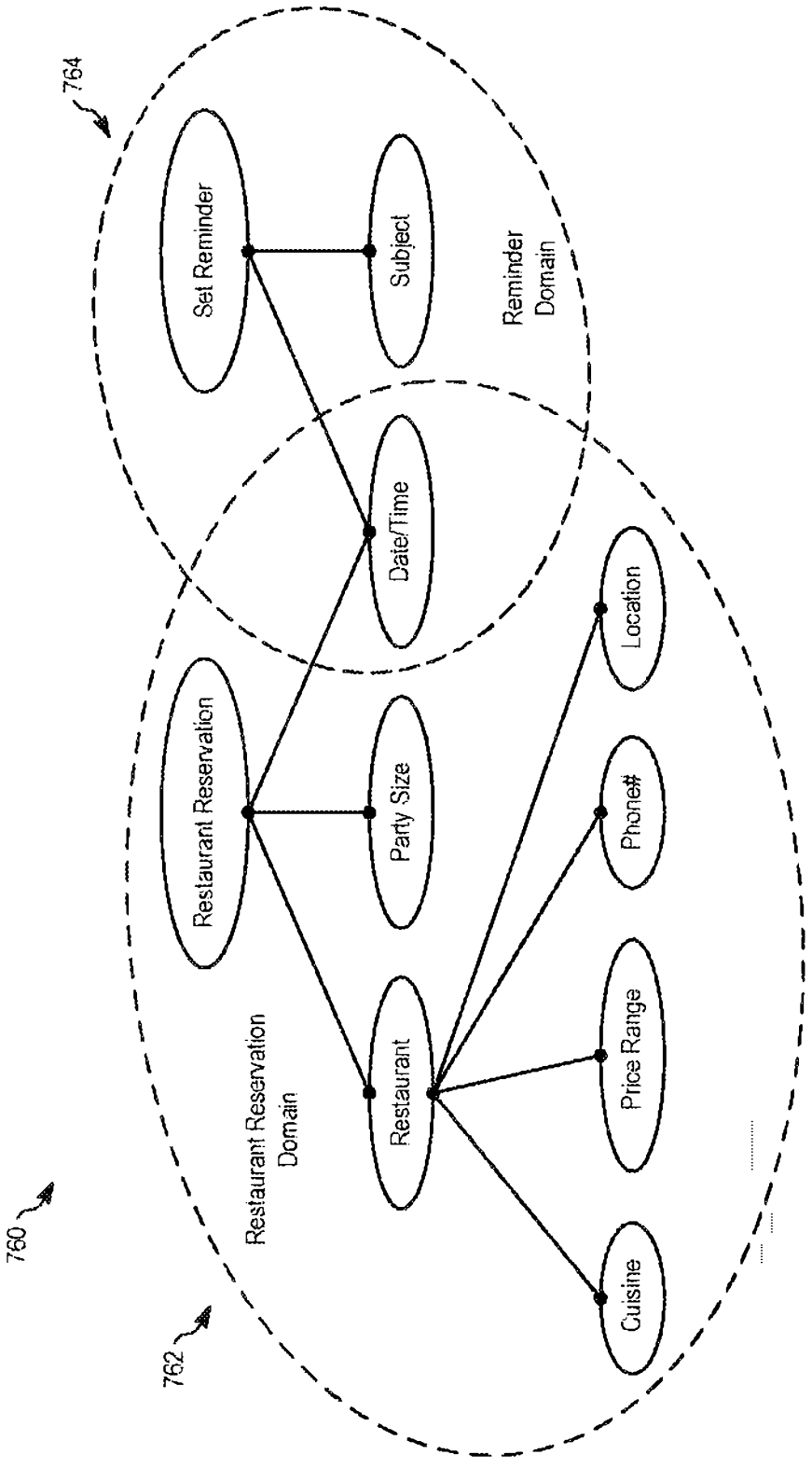


FIG. 7C

DK 2017 70434 A1

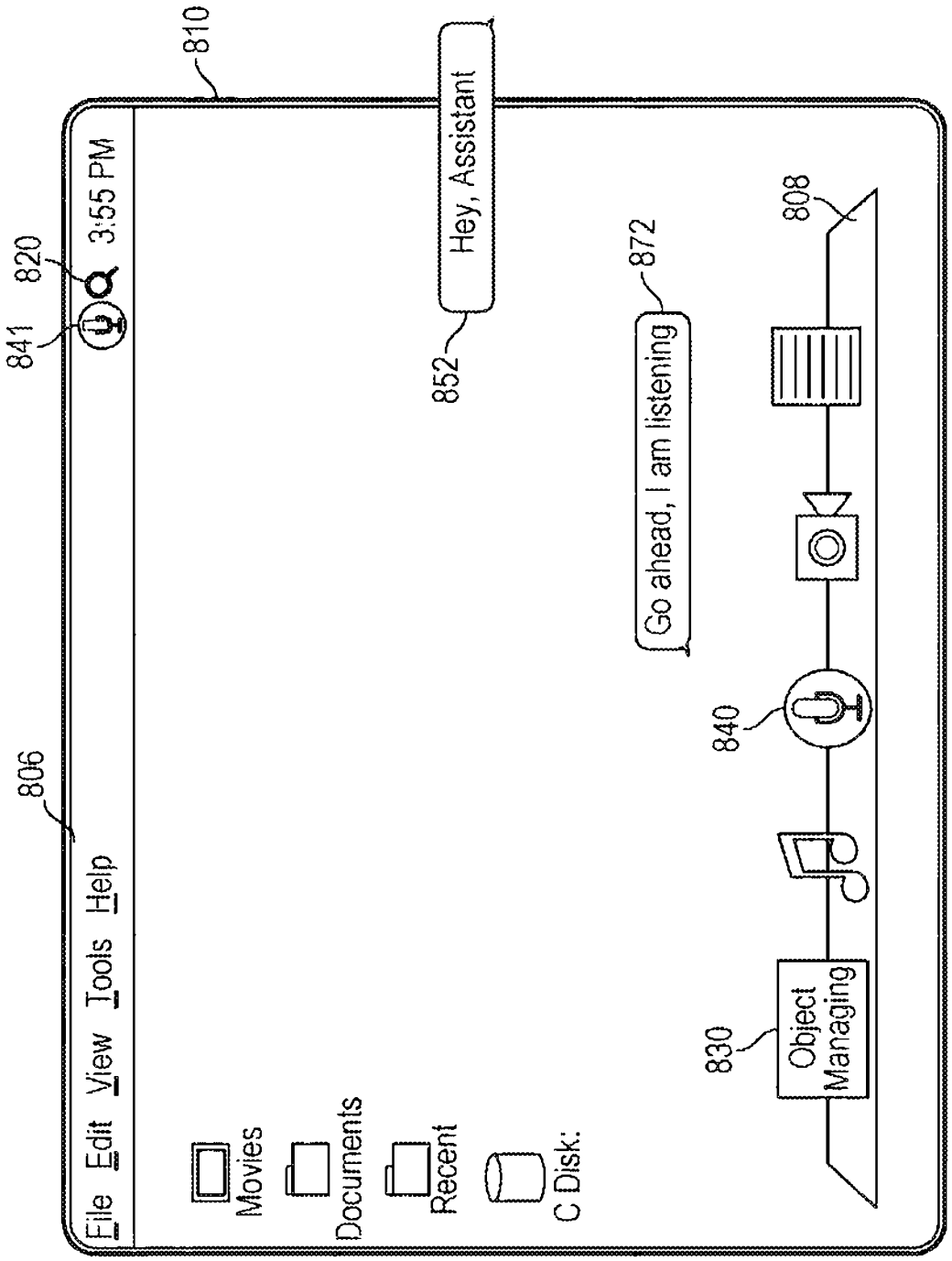


FIG. 8A

DK 2017 70434 A1

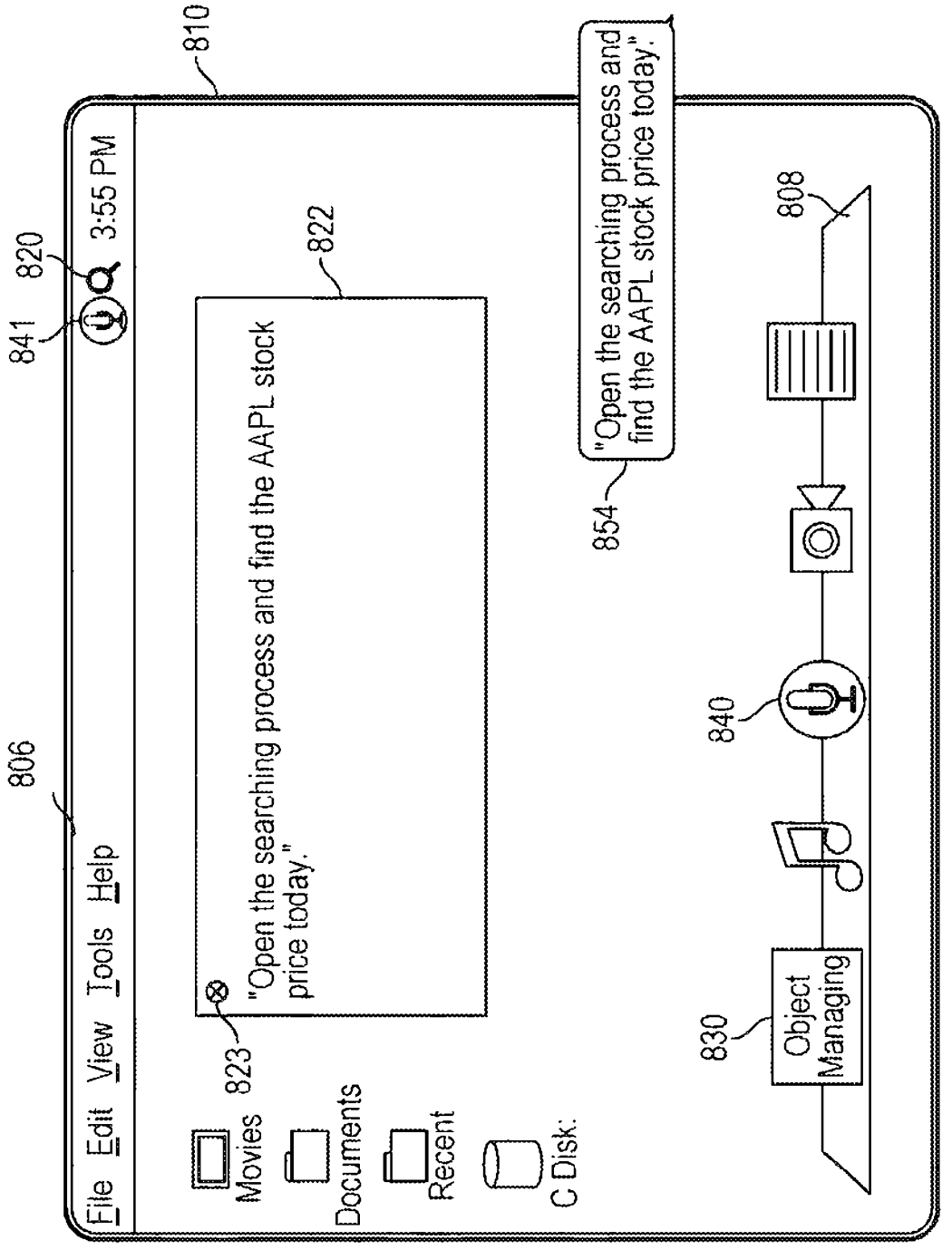


FIG. 8B

DK 2017 70434 A1

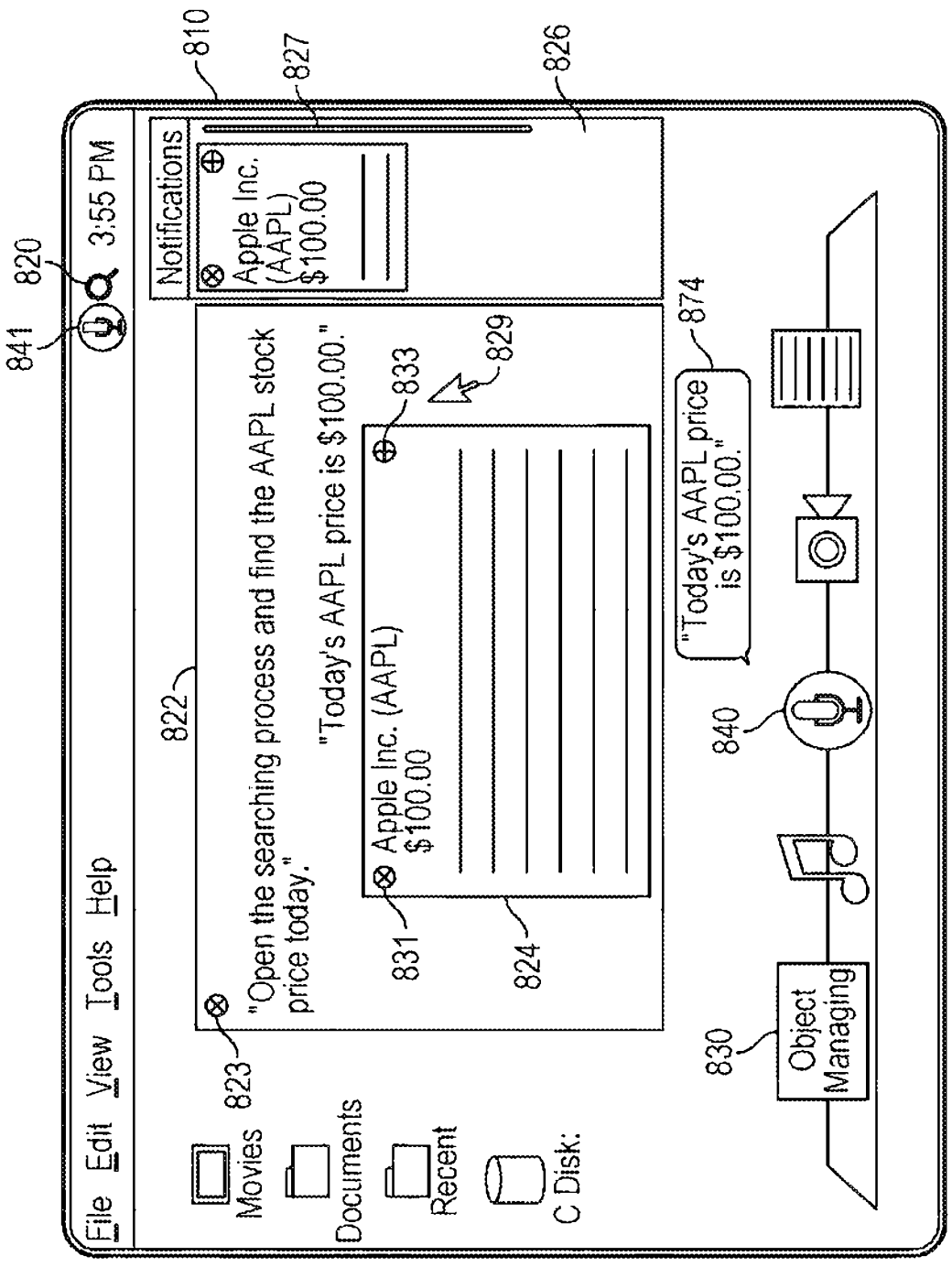


FIG. 8C

DK 2017 70434 A1

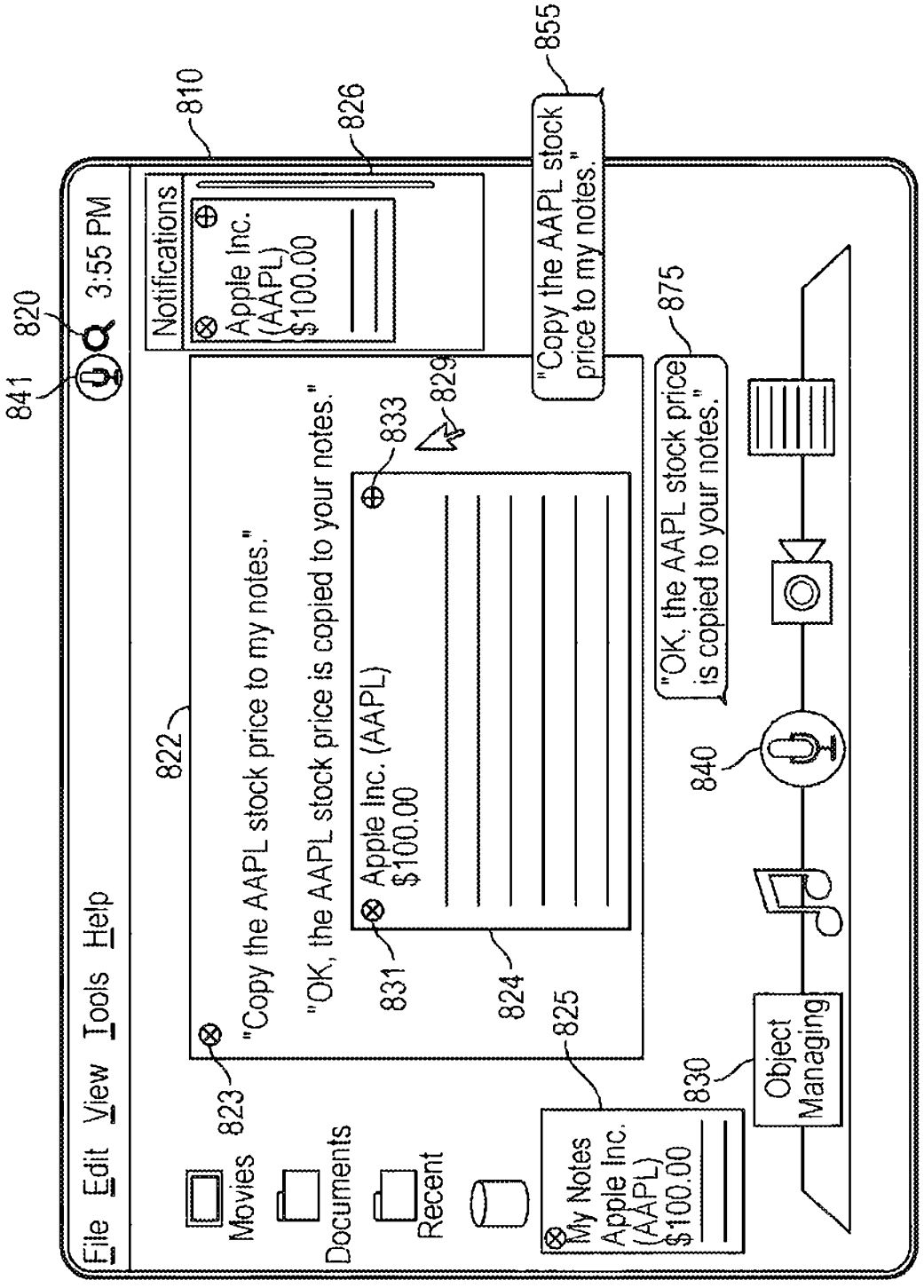


FIG. 8D

DK 2017 70434 A1

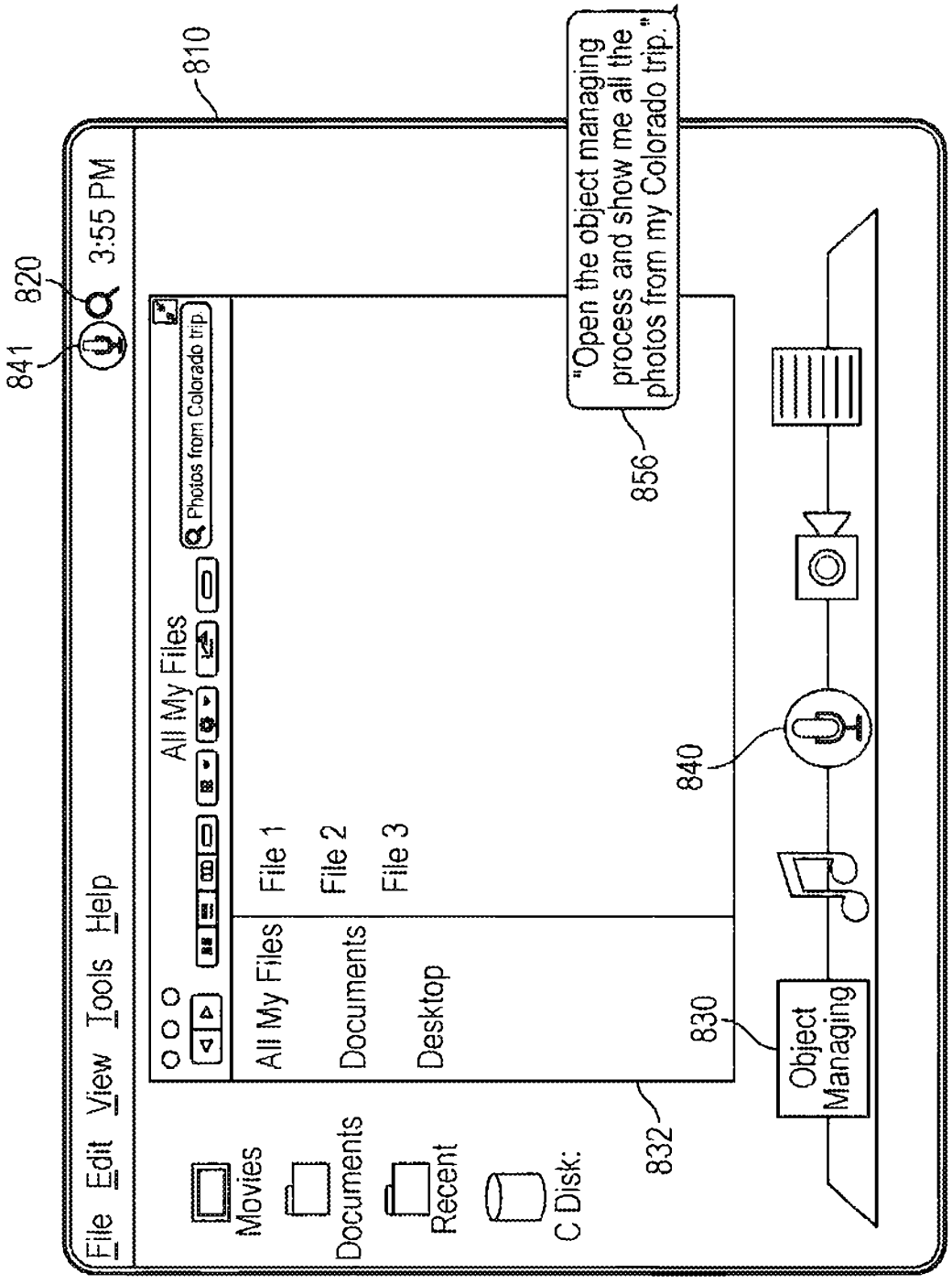


FIG. 8E

DK 2017 70434 A1

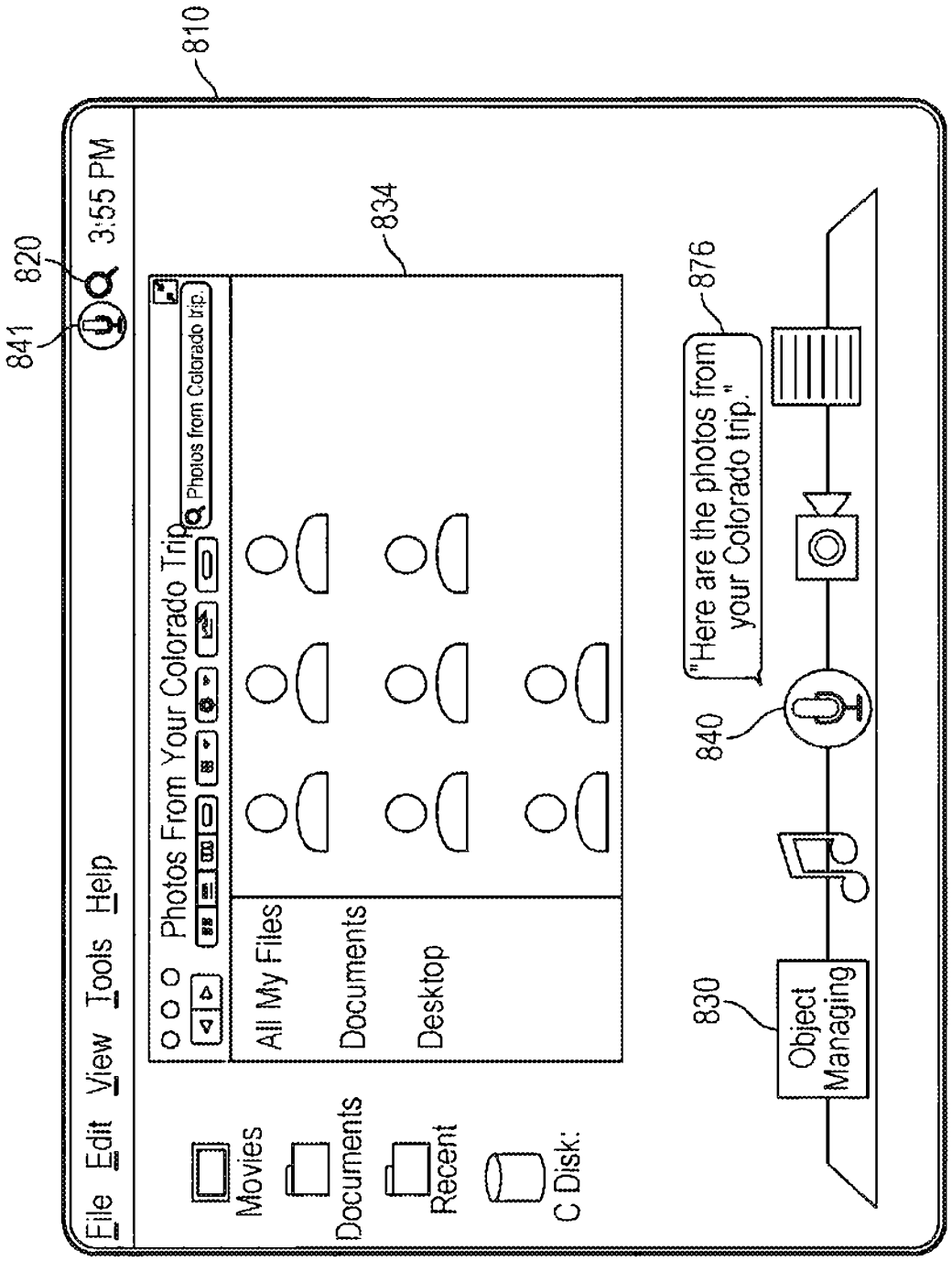


FIG. 8F

DK 2017 70434 A1

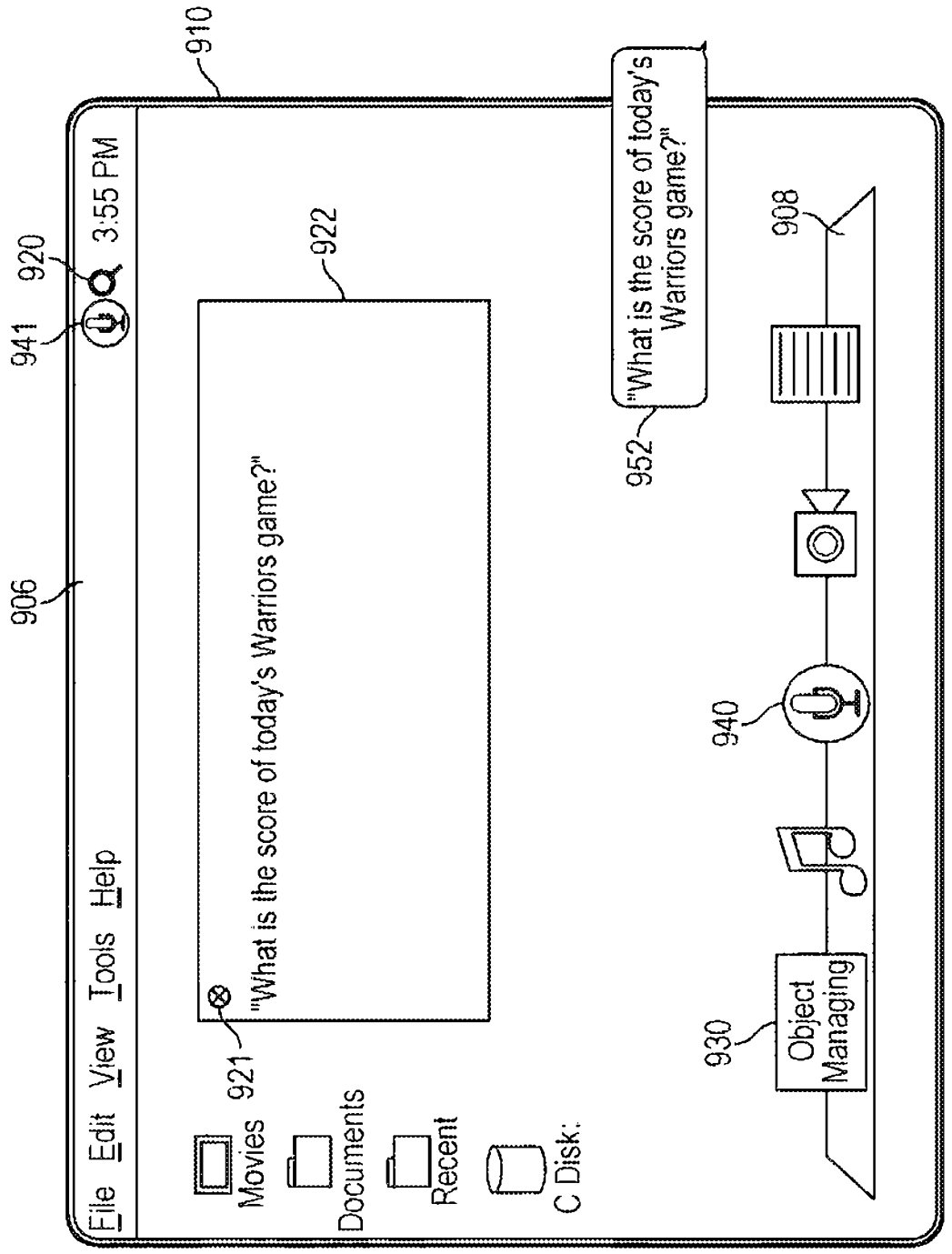


FIG. 9A

DK 2017 70434 A1

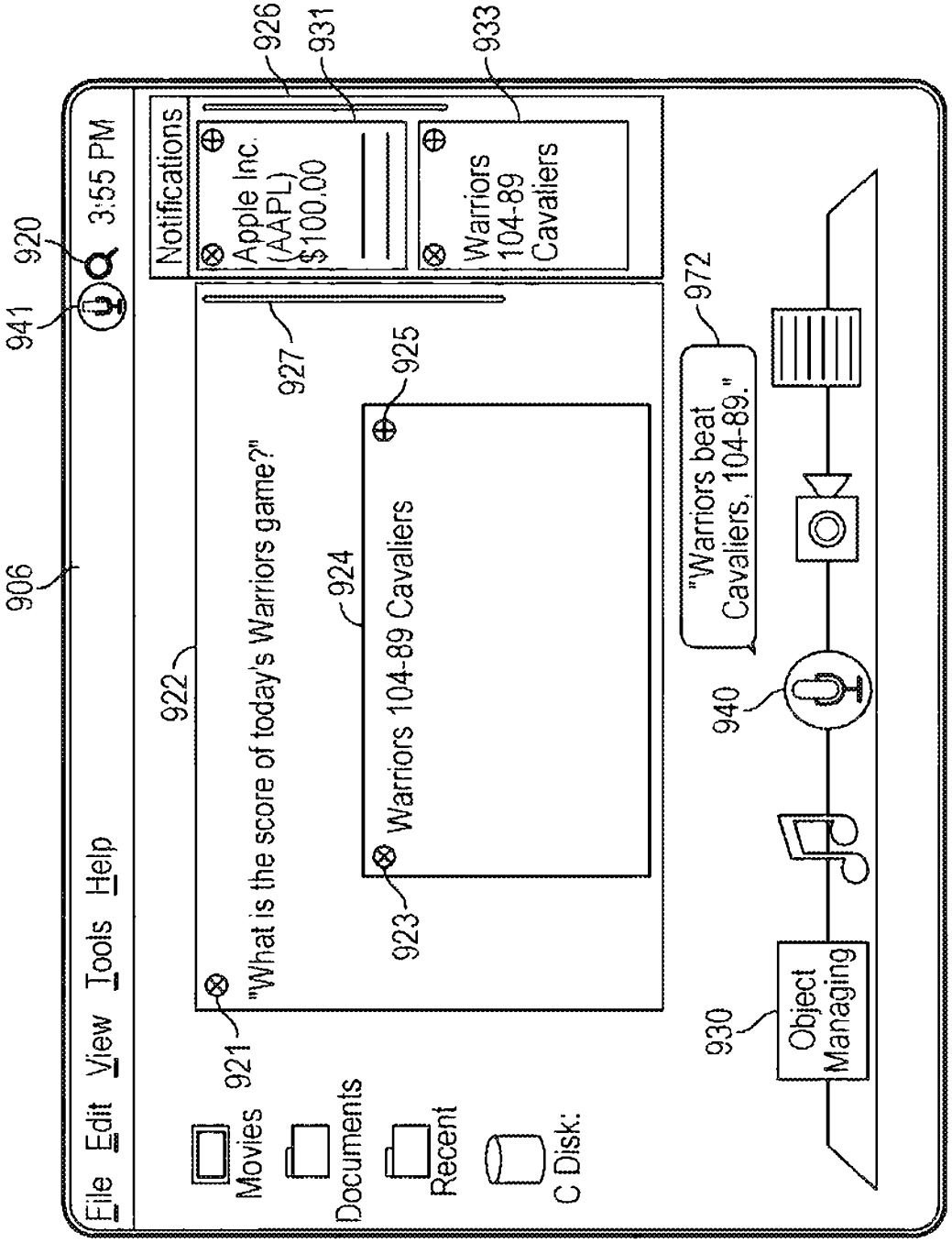


FIG. 9B

DK 2017 70434 A1

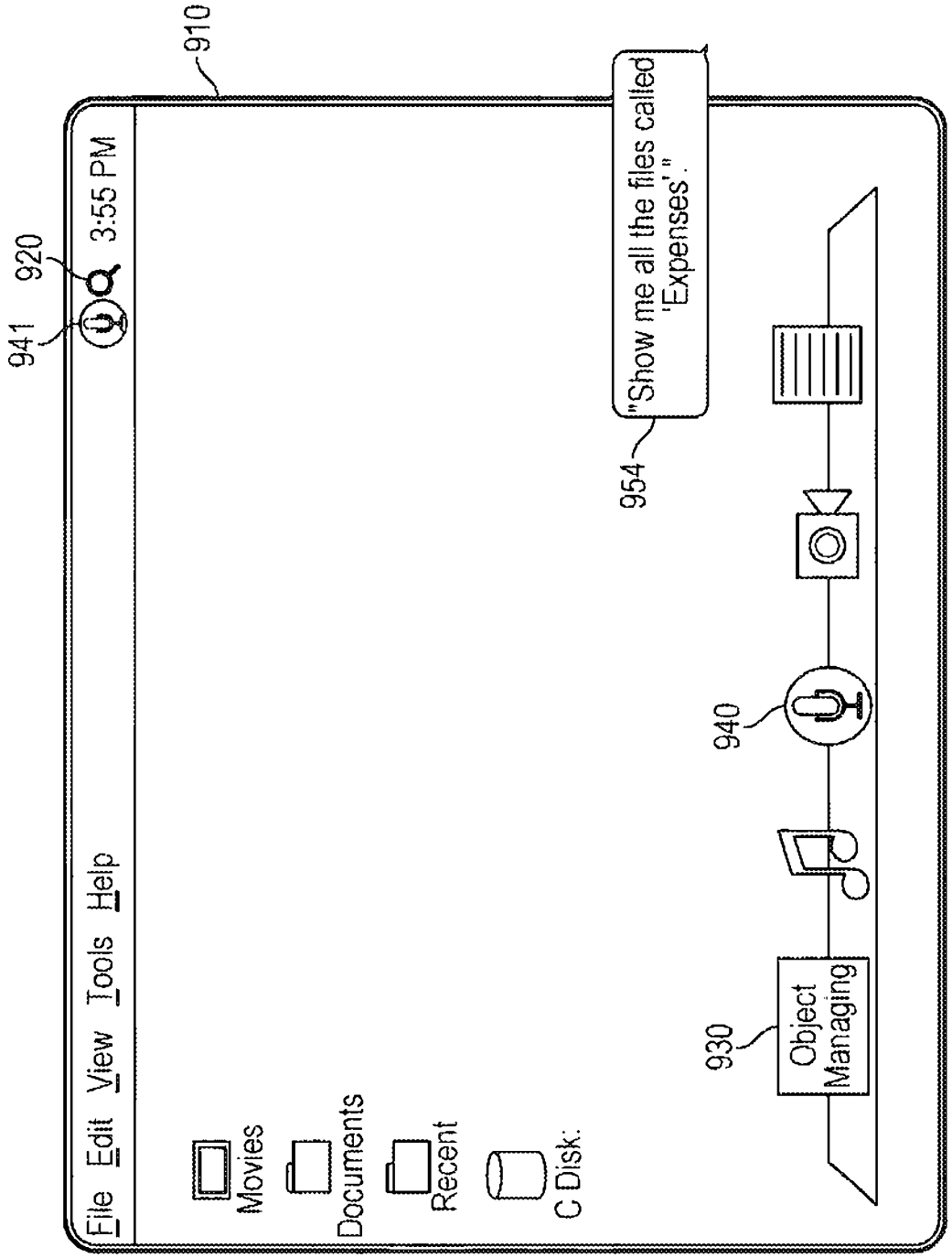


FIG. 9C

DK 2017 70434 A1

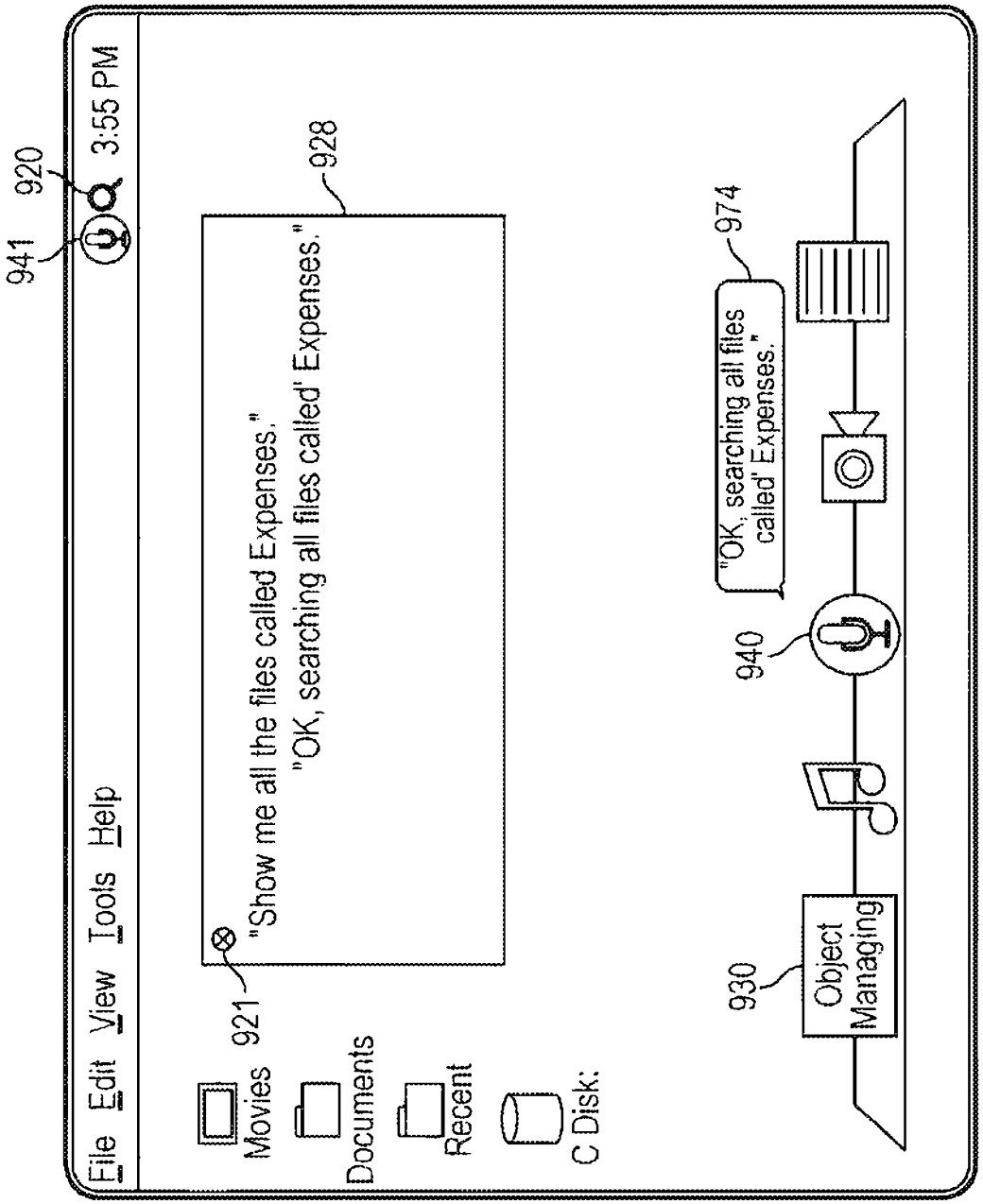


FIG. 9D

DK 2017 70434 A1

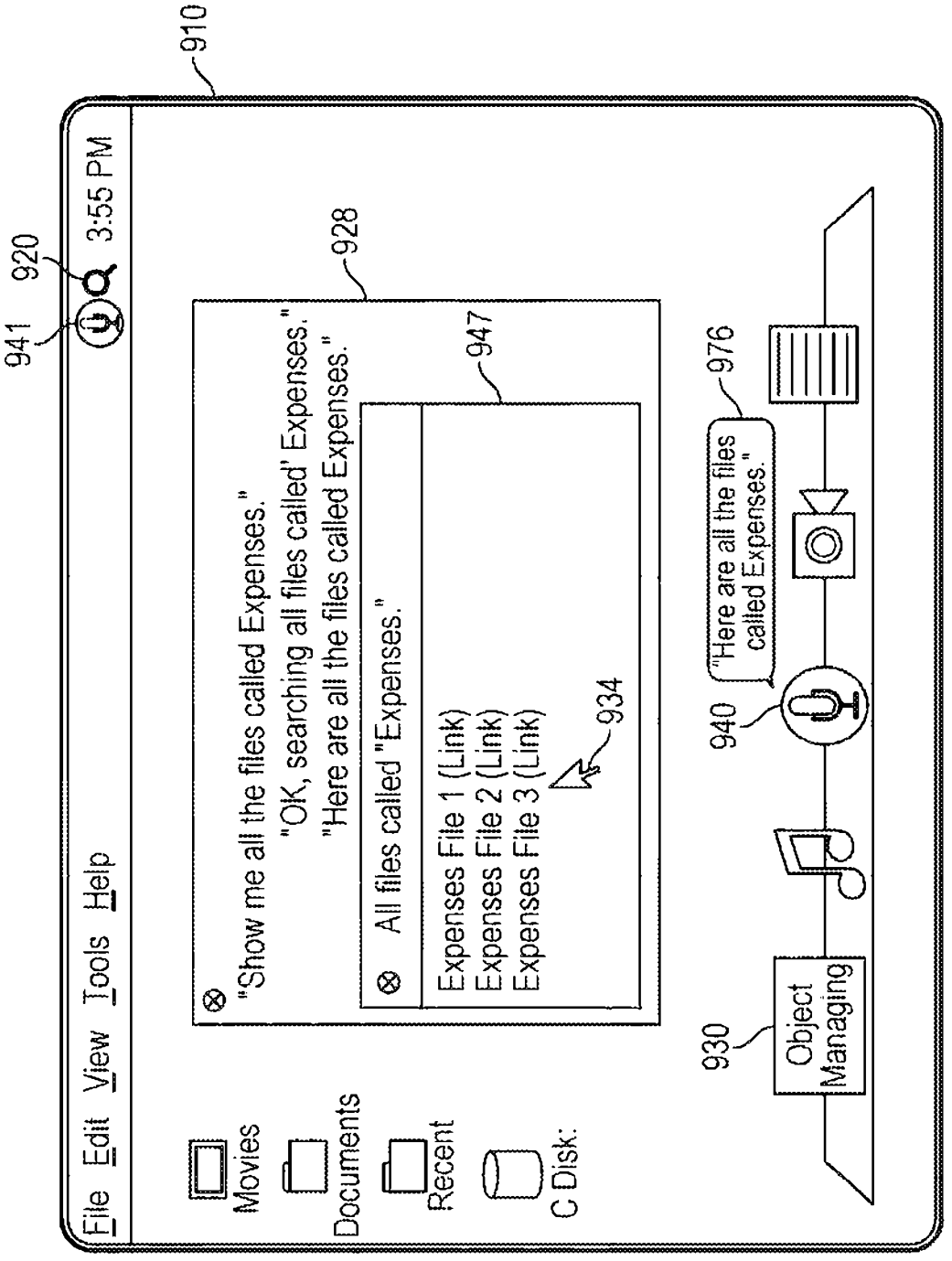


FIG. 9E

DK 2017 70434 A1

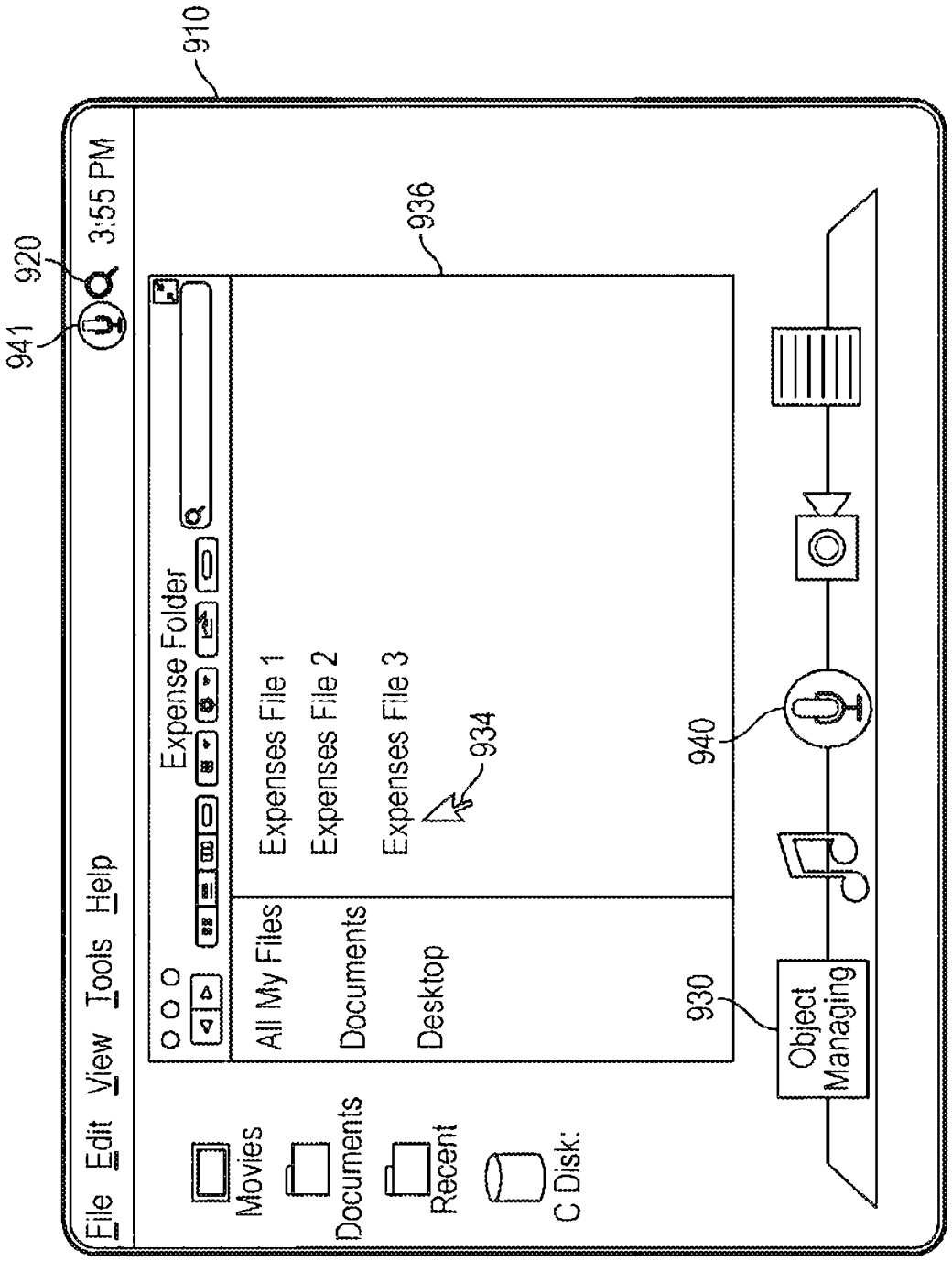


FIG. 9F

DK 2017 70434 A1

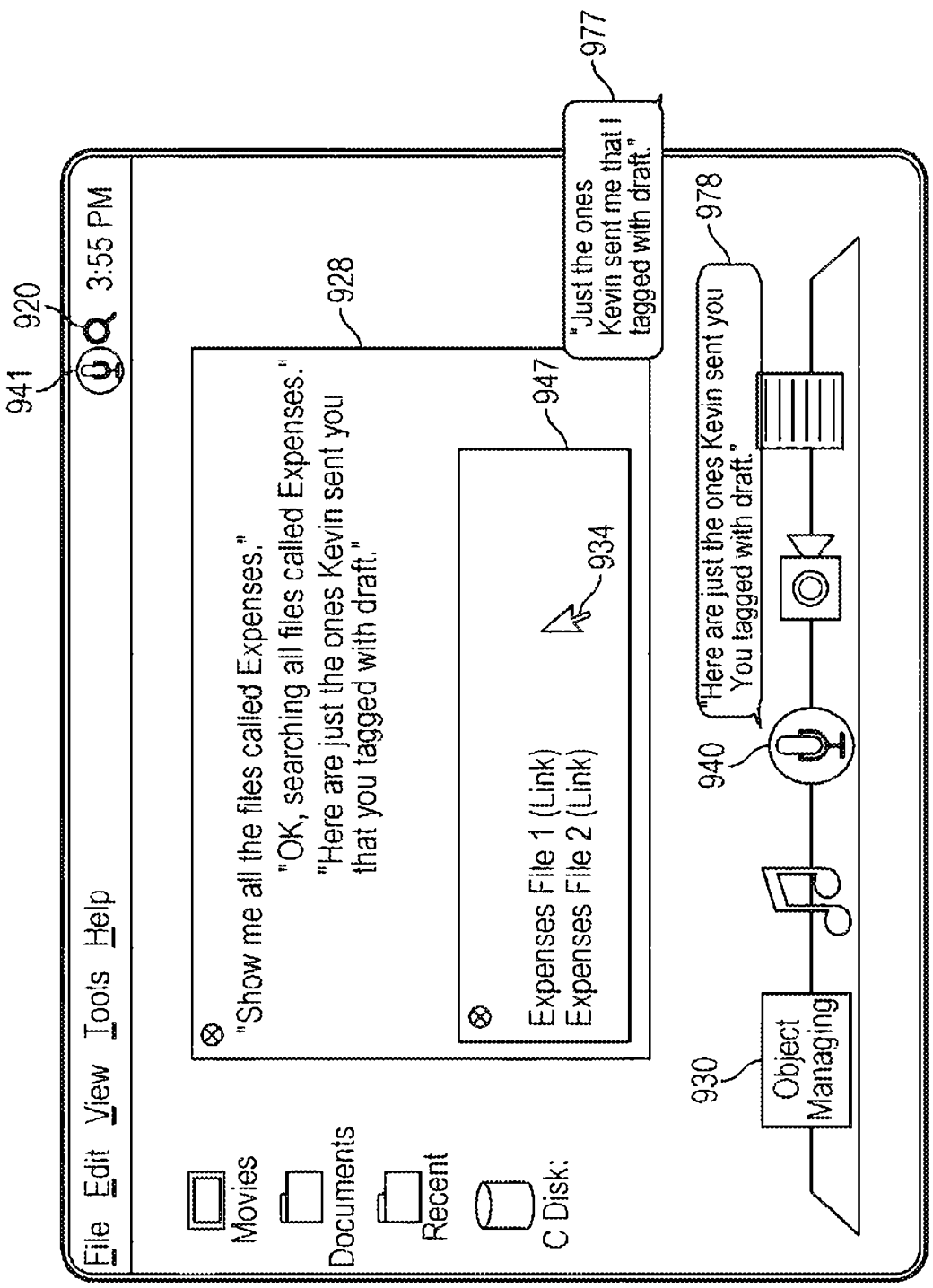


FIG. 9G

DK 2017 70434 A1

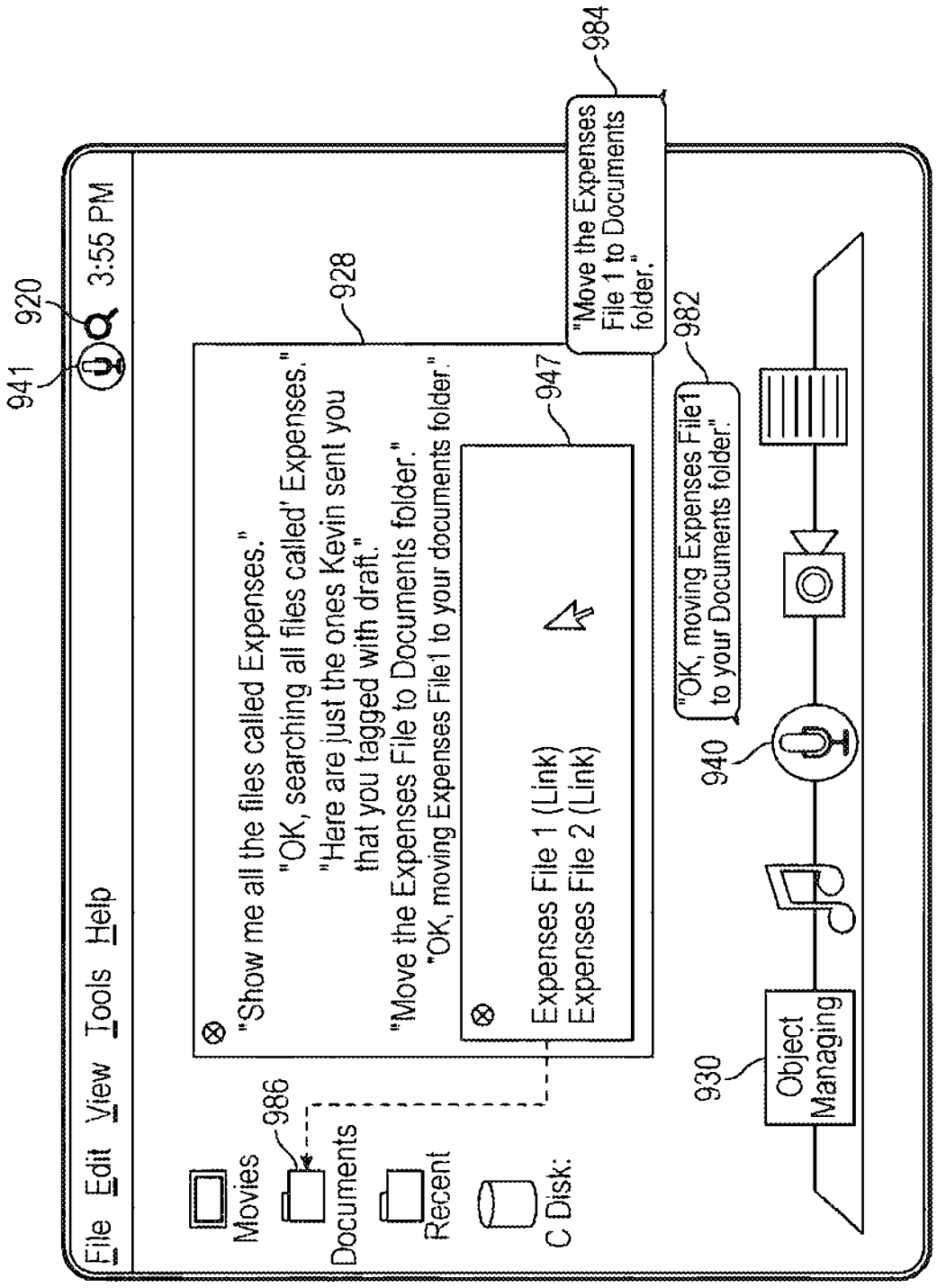


FIG. 9H

DK 2017 70434 A1

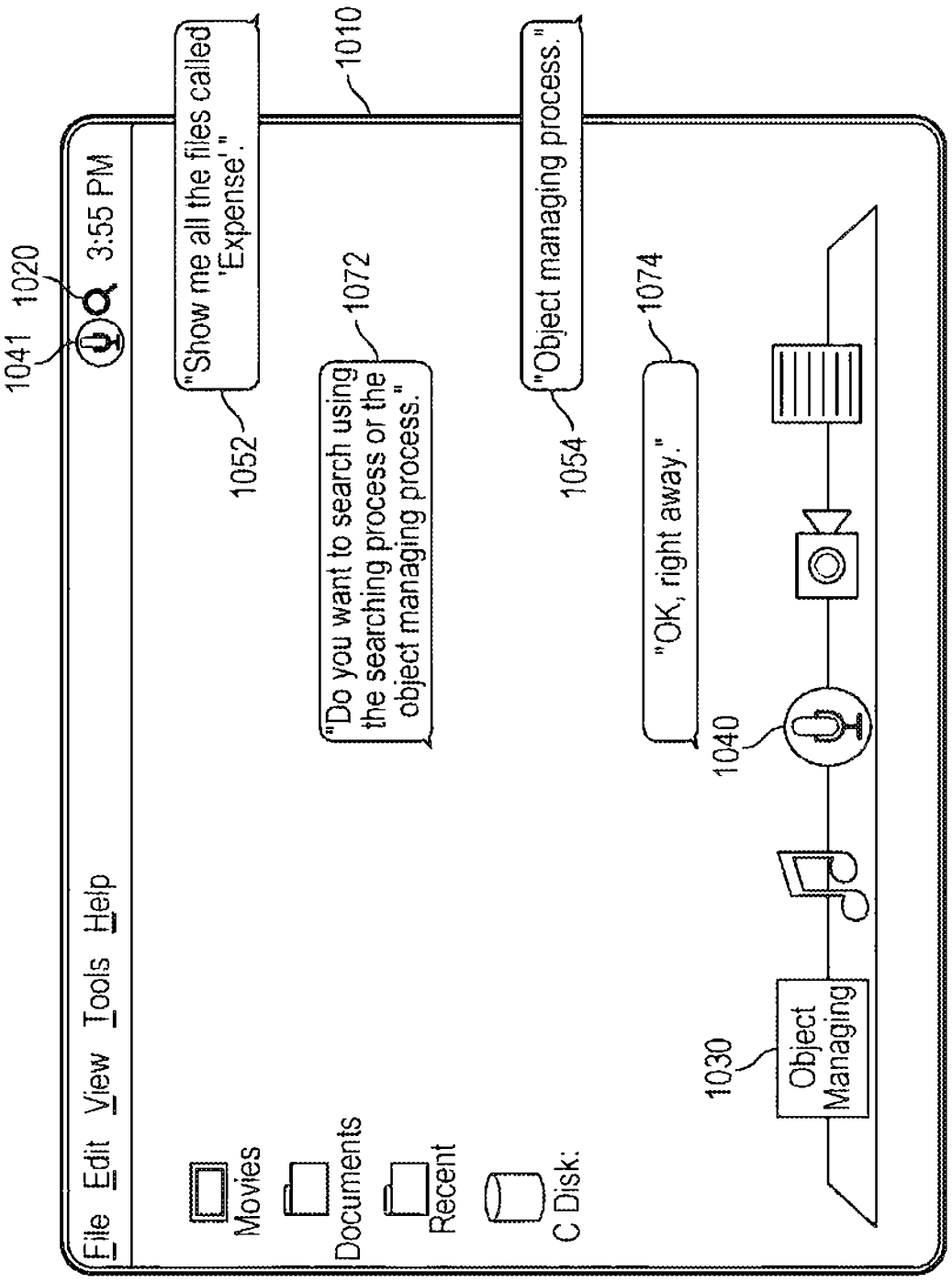


FIG. 10A

DK 2017 70434 A1

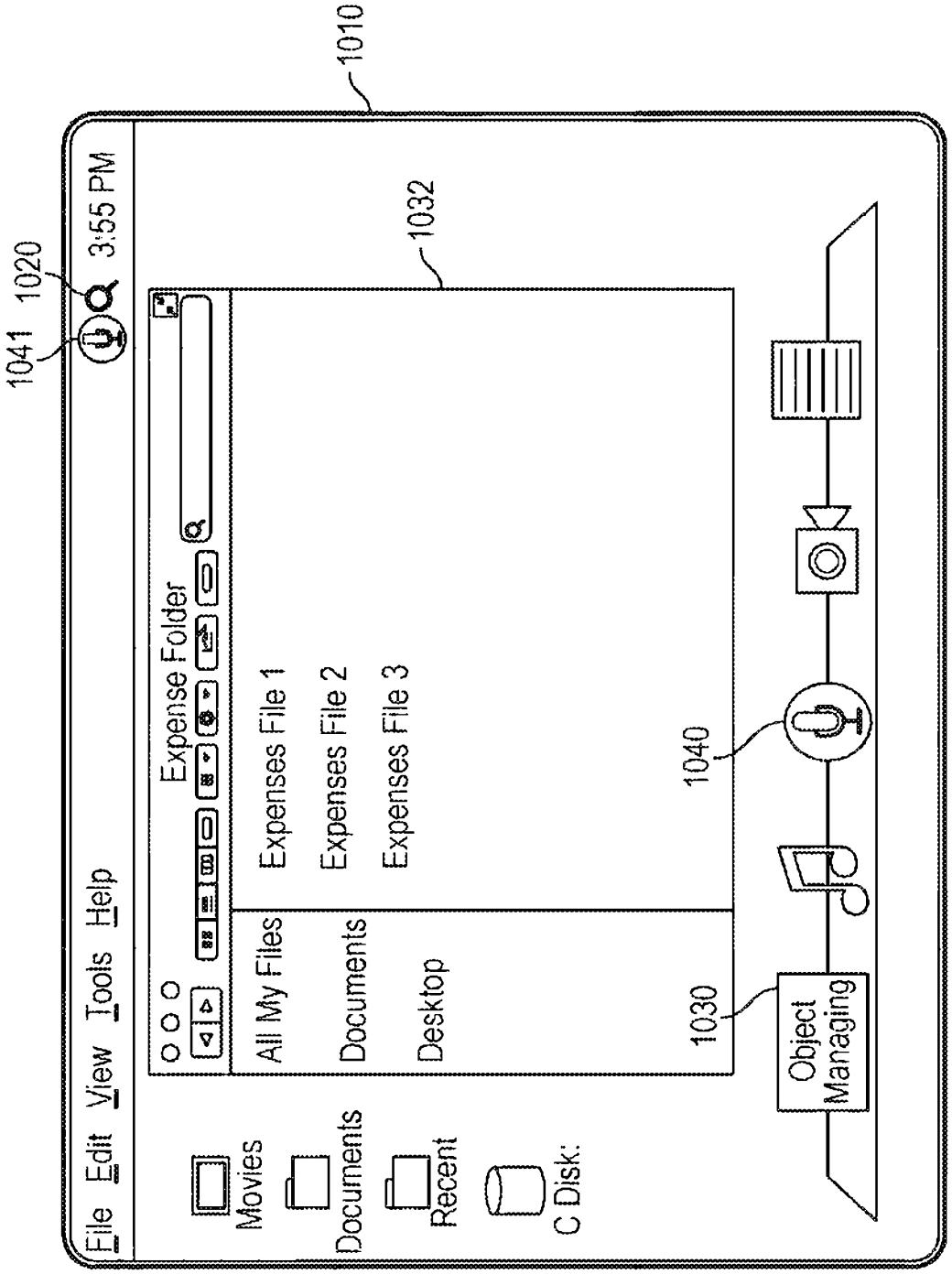


FIG. 10B

DK 2017 70434 A1

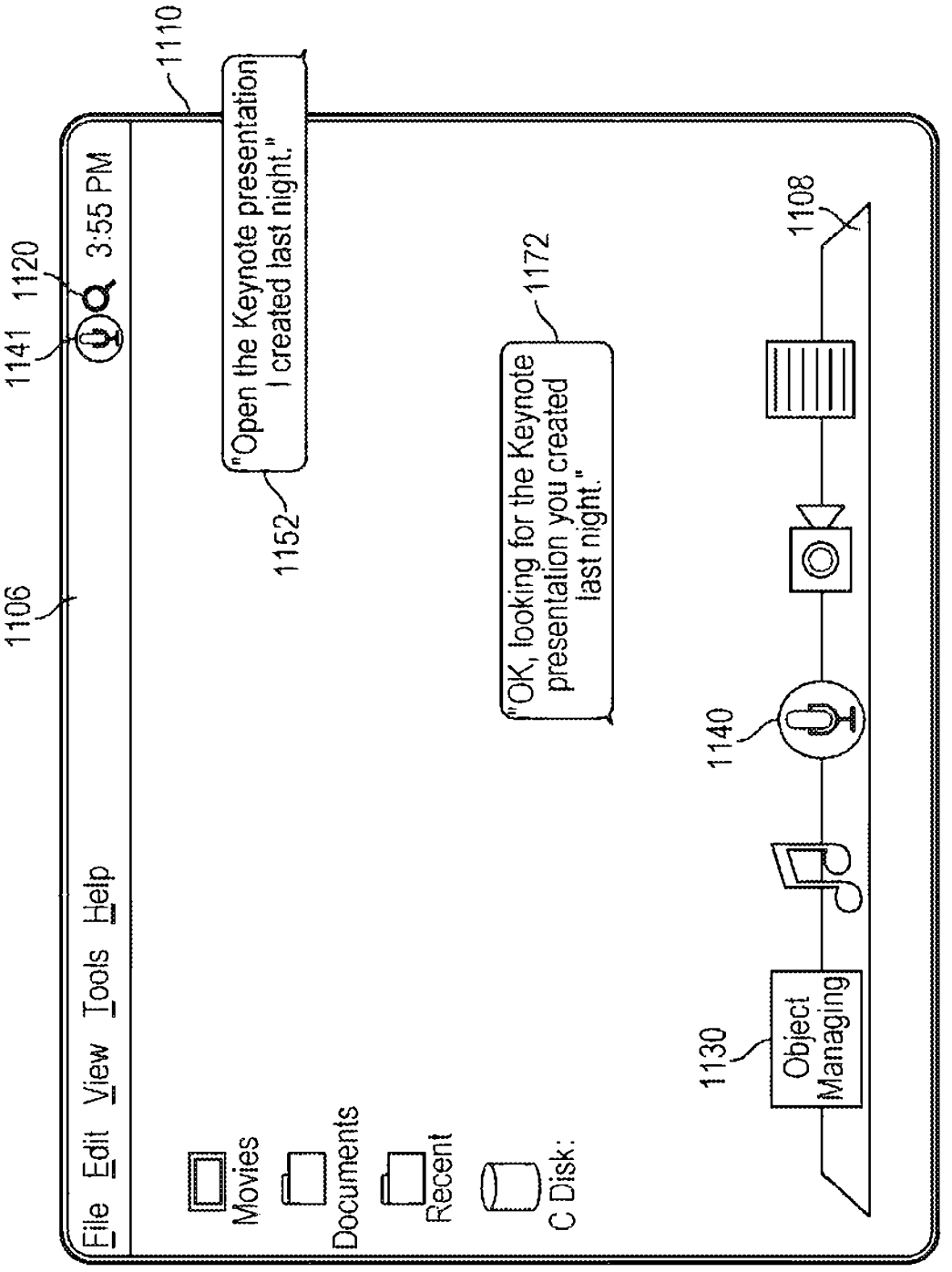


FIG. 11A

DK 2017 70434 A1

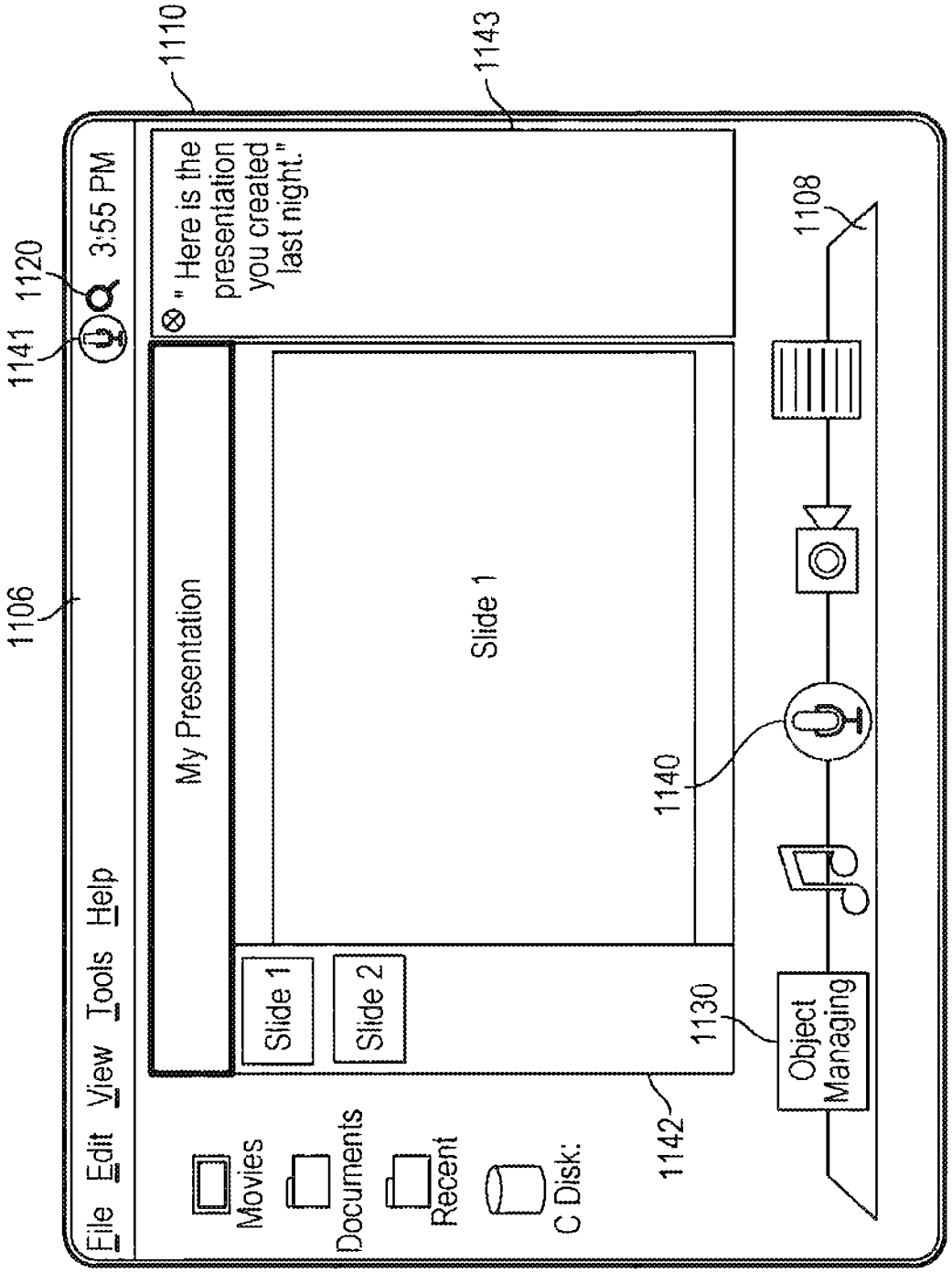


FIG. 11B

DK 2017 70434 A1

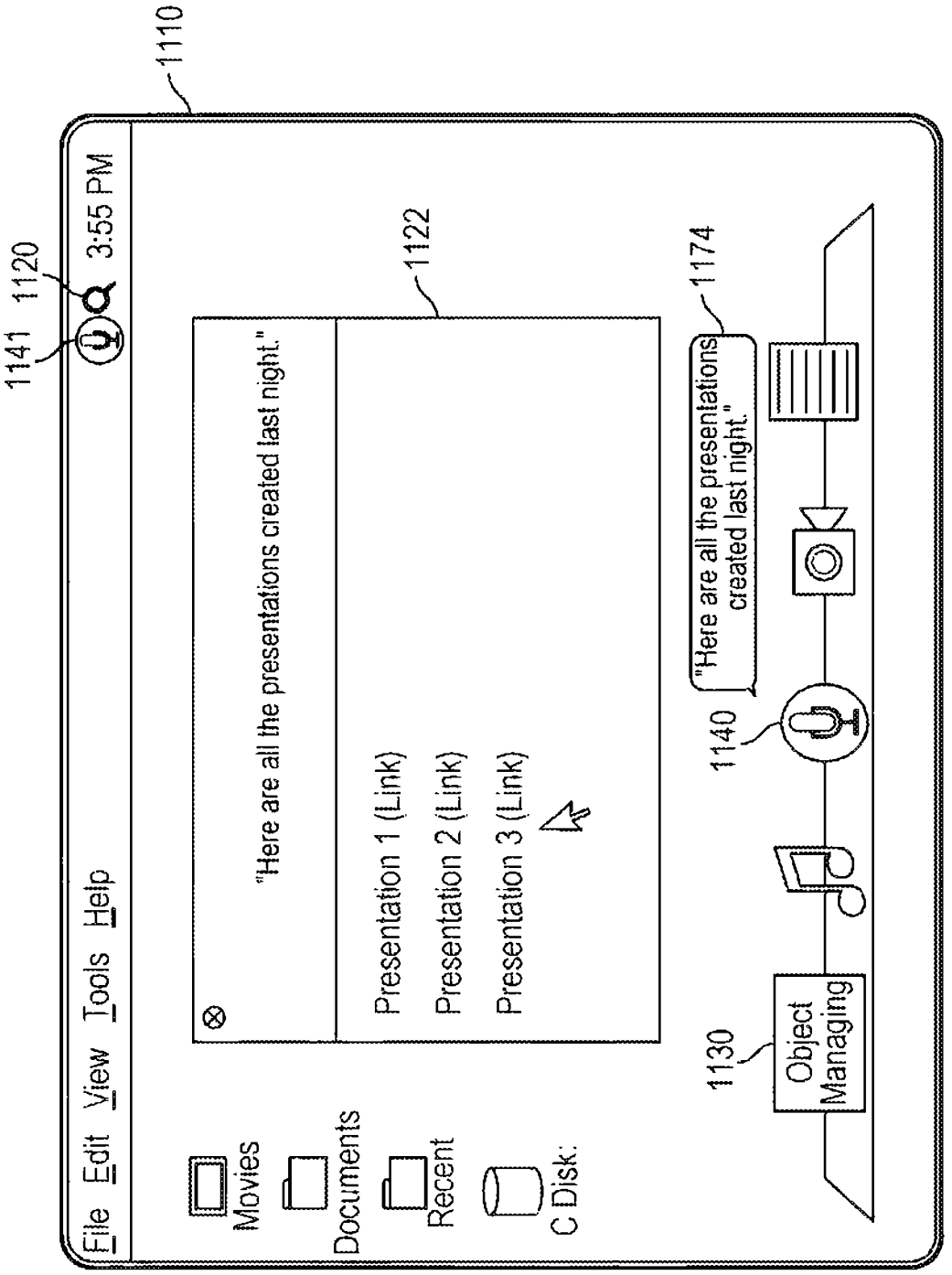


FIG. 11C

DK 2017 70434 A1

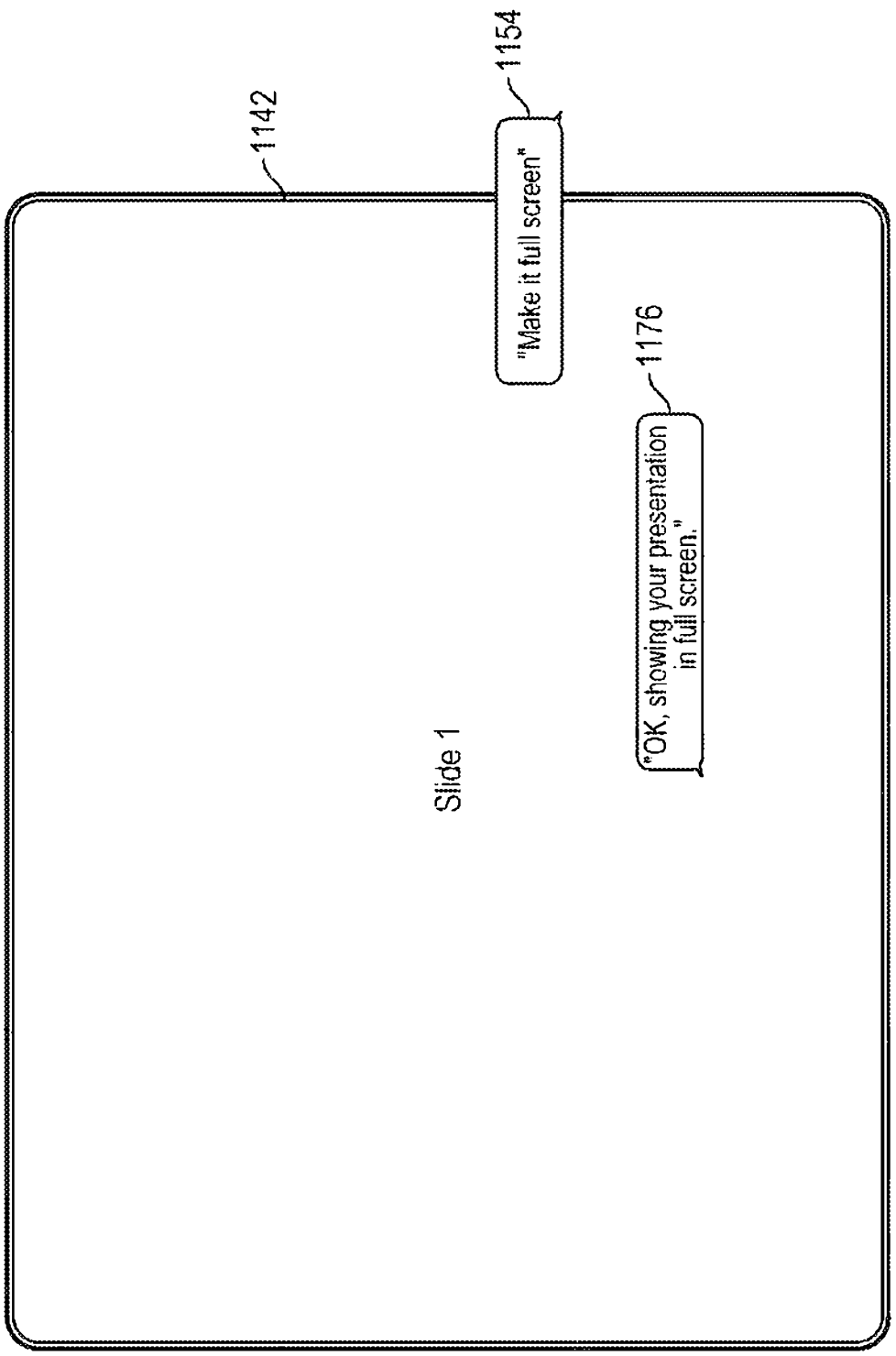


FIG. 11D

DK 2017 70434 A1

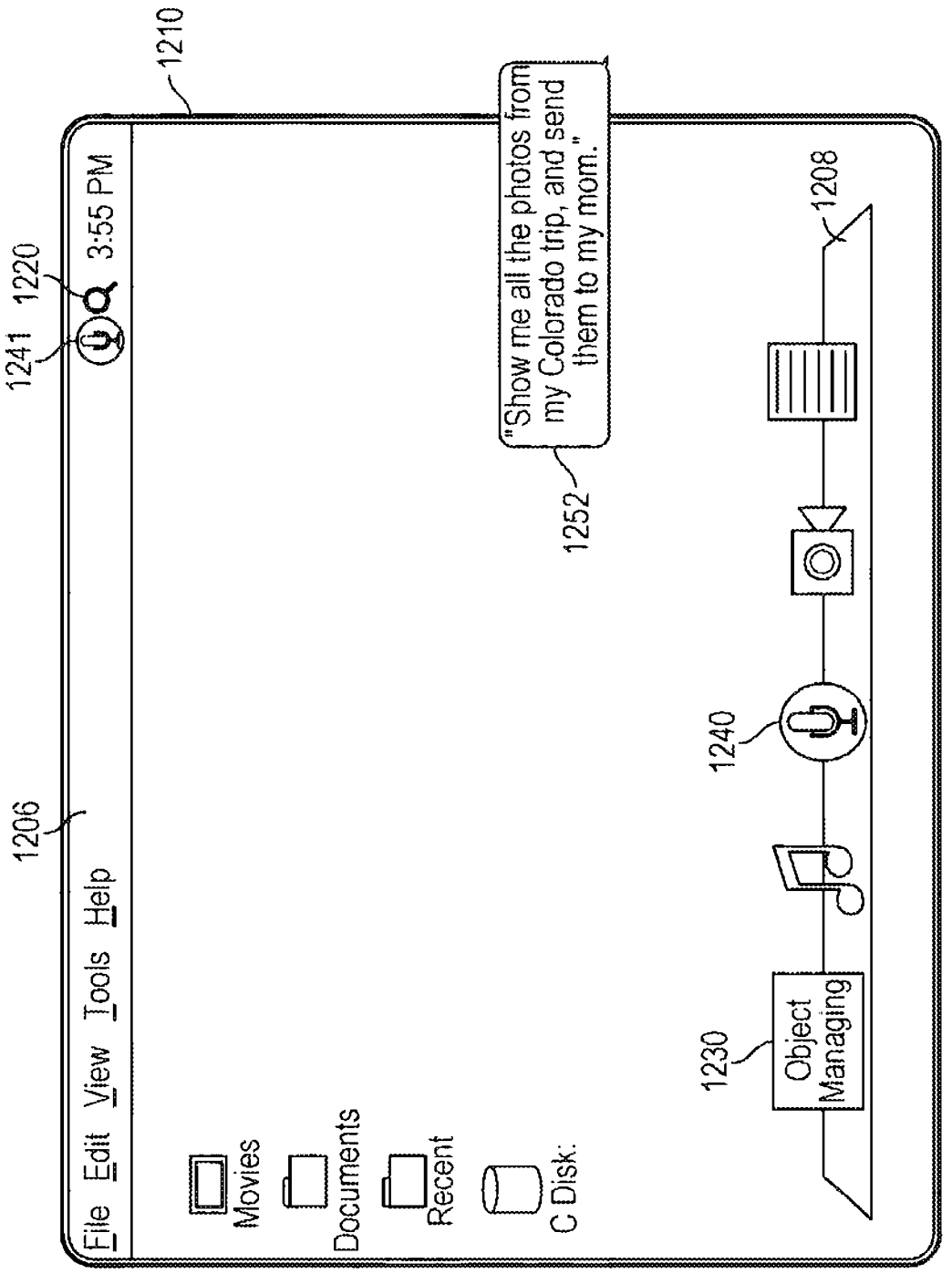


FIG. 12A

DK 2017 70434 A1

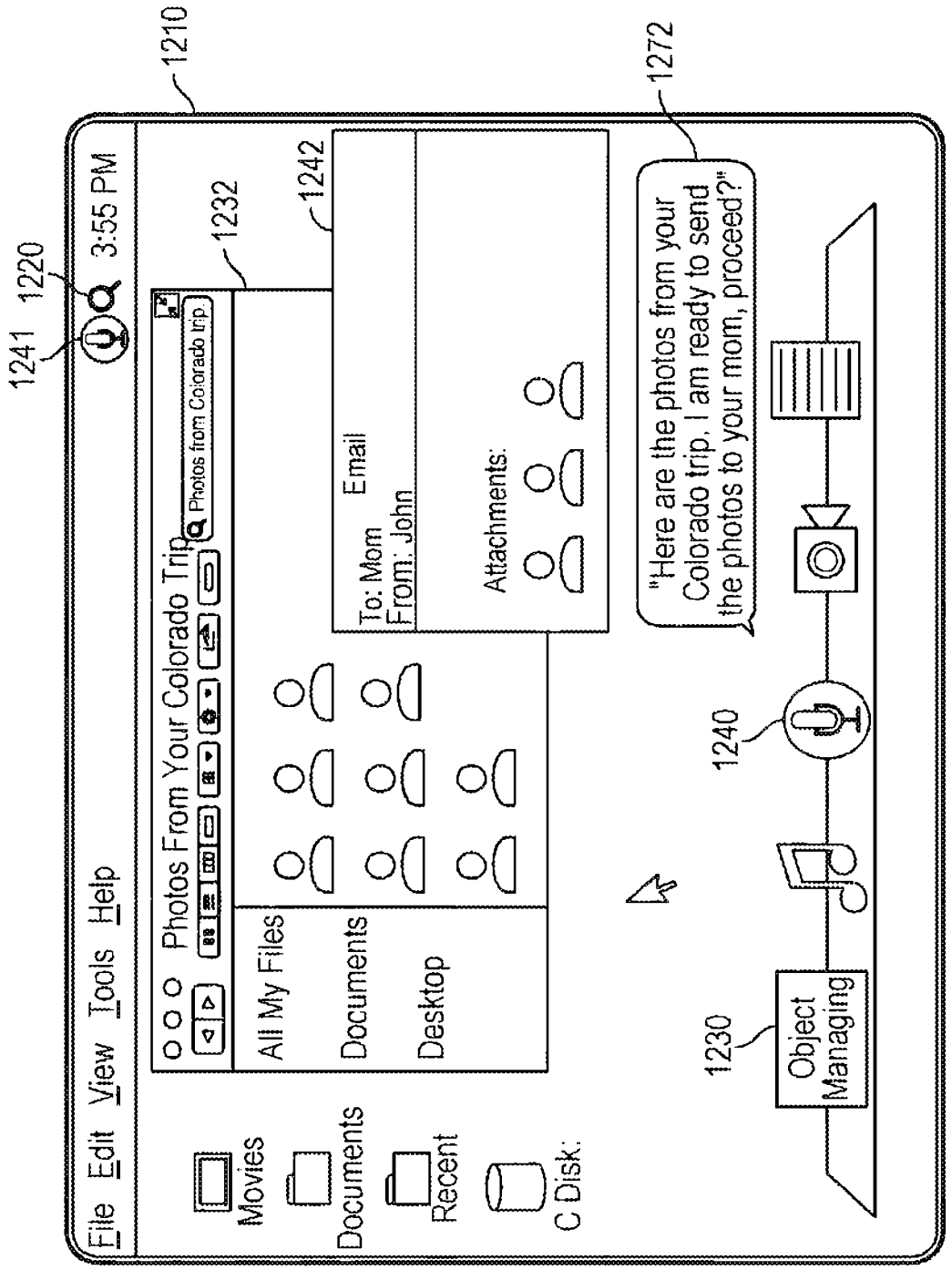


FIG. 12B

DK 2017 70434 A1

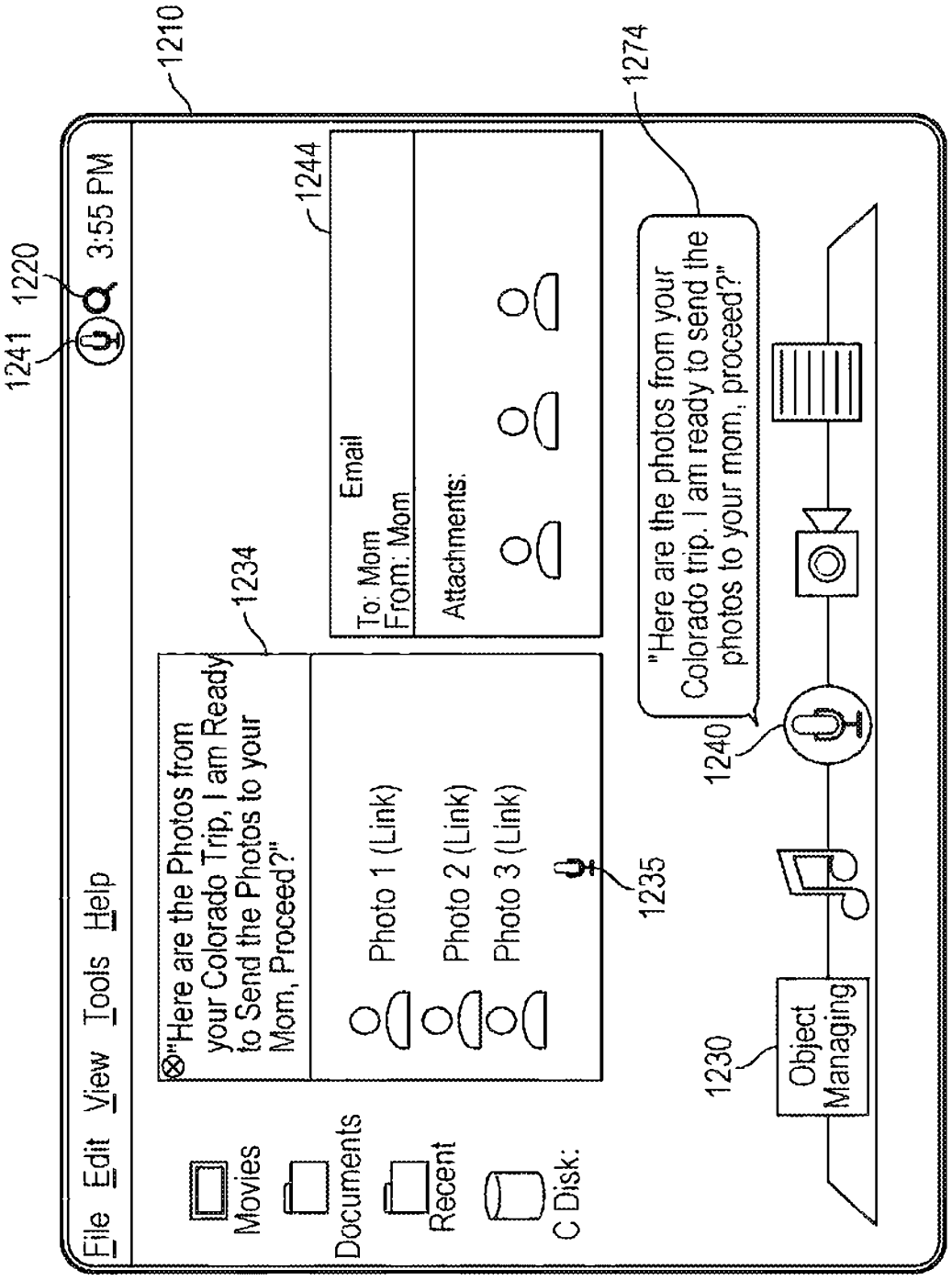


FIG. 12C

DK 2017 70434 A1

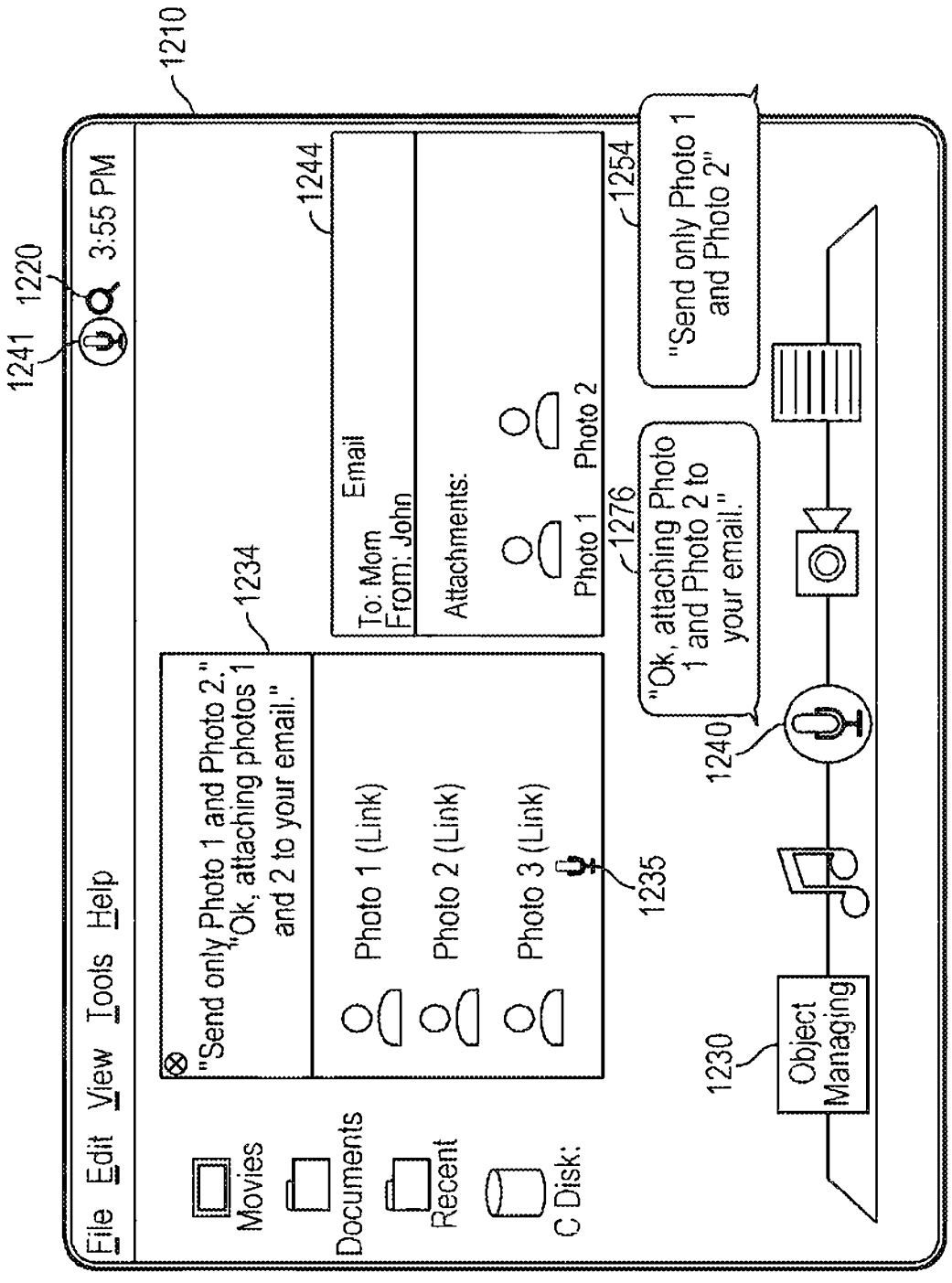


FIG. 12D

DK 2017 70434 A1

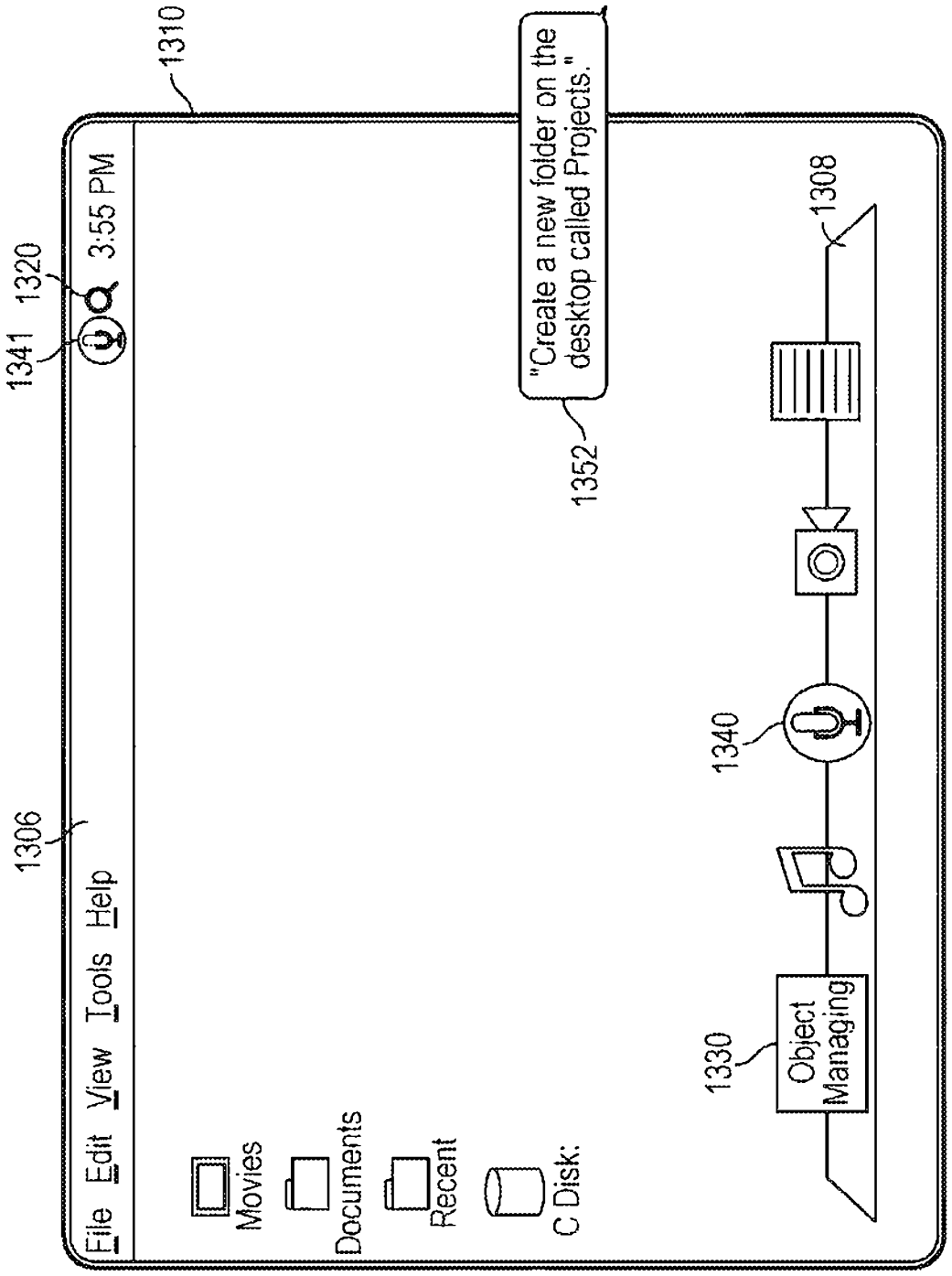


FIG. 13A

DK 2017 70434 A1

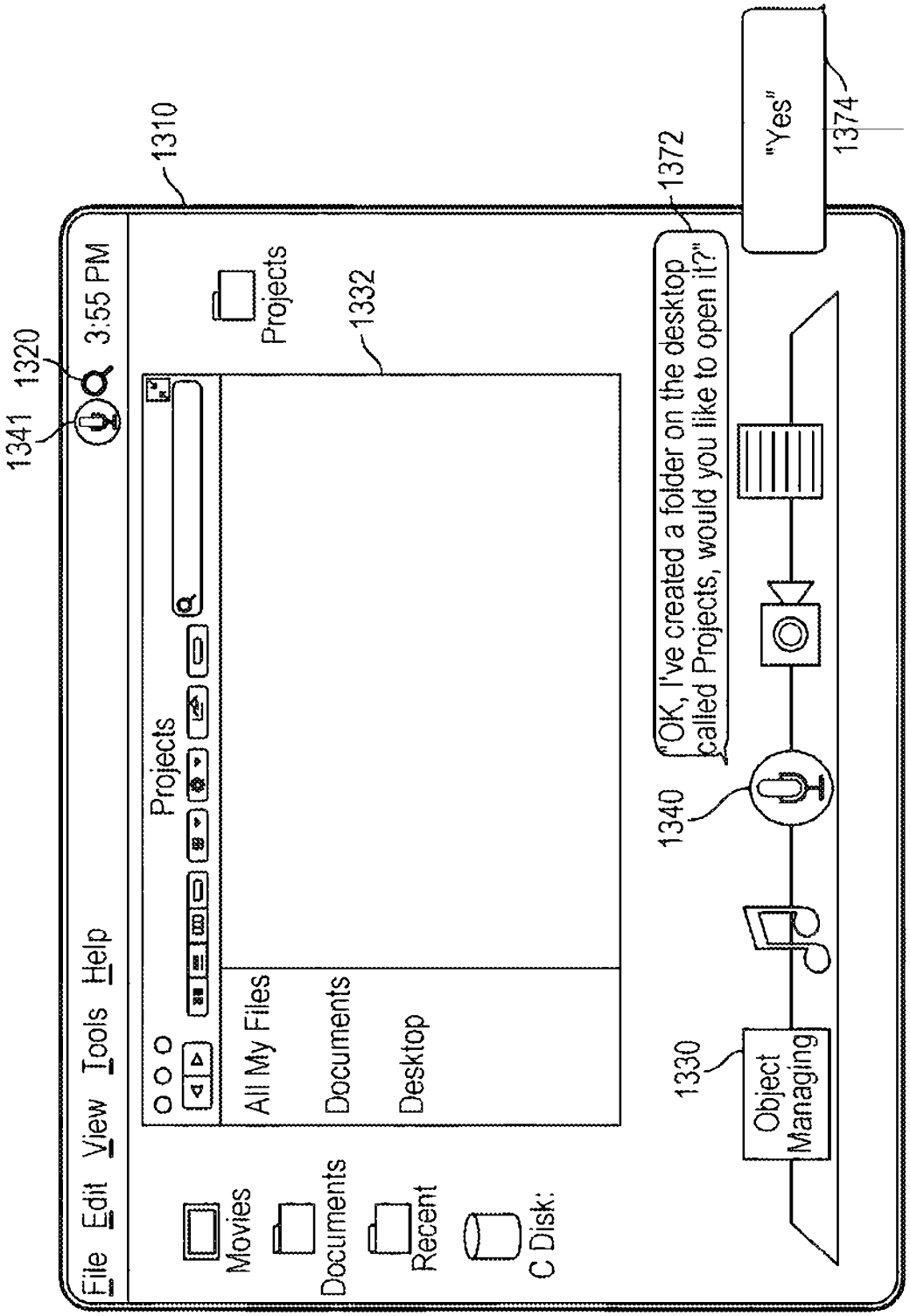


FIG. 13B

DK 2017 70434 A1

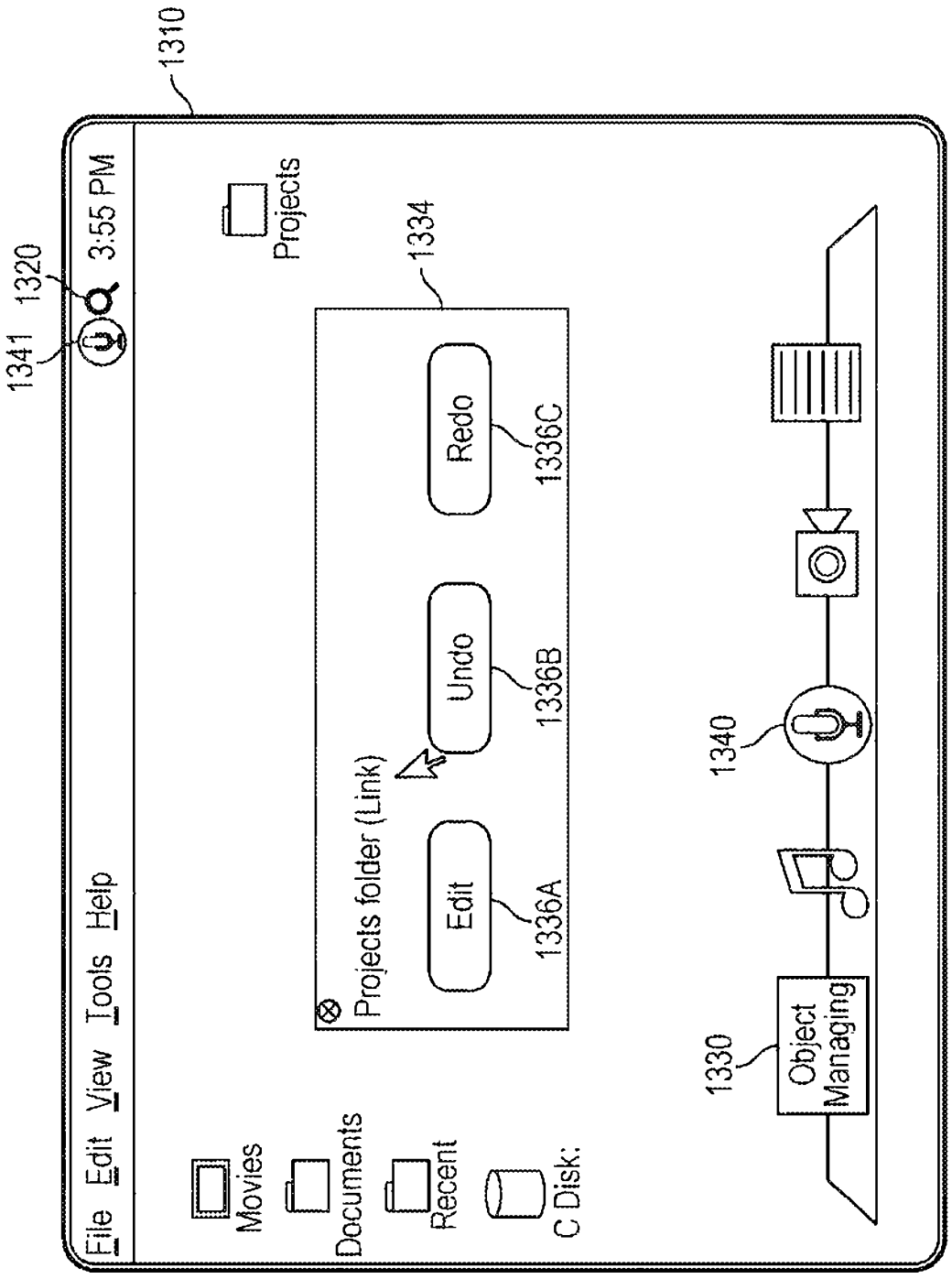


FIG. 13C

DK 2017 70434 A1

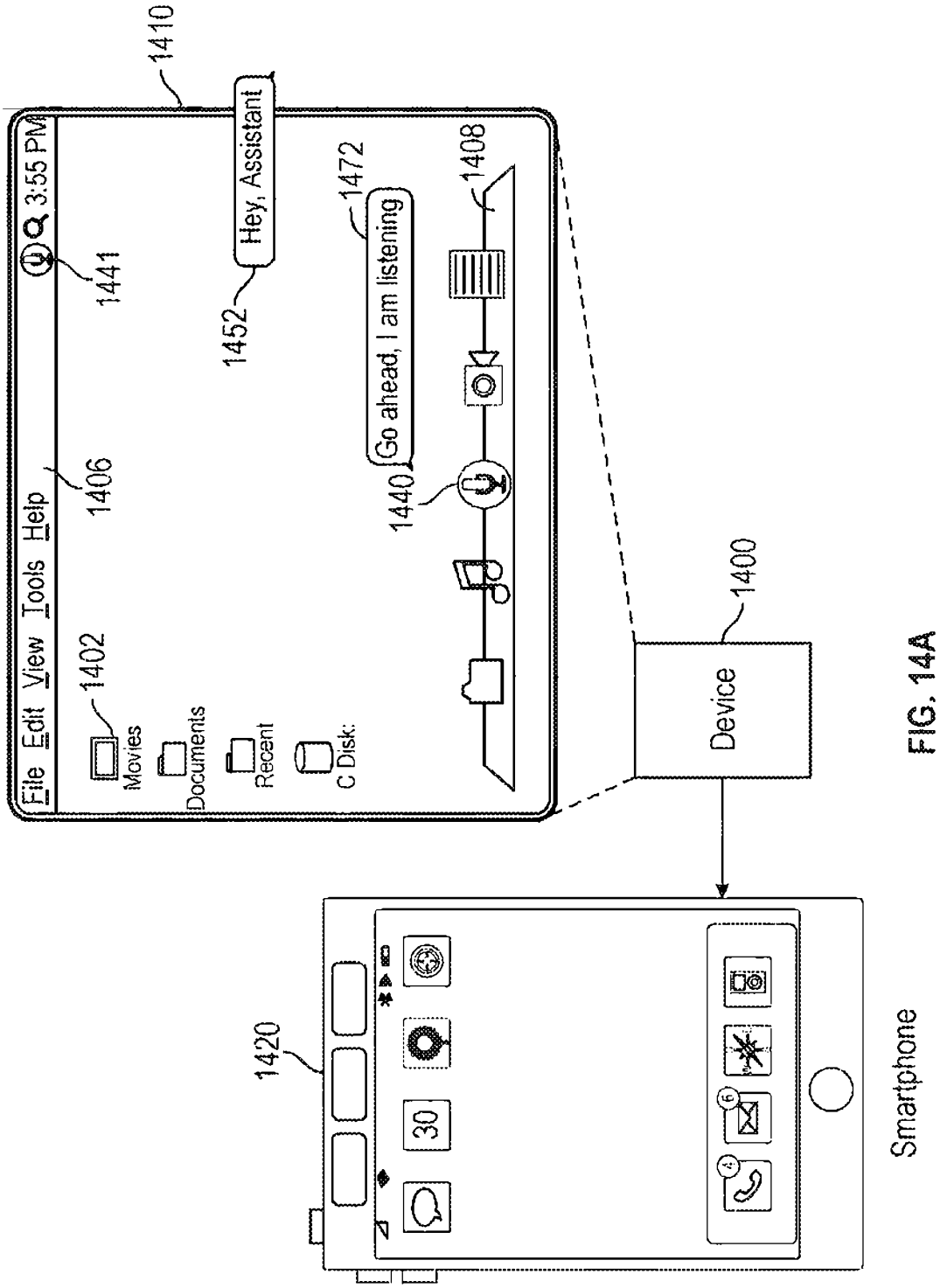


FIG. 14A

DK 2017 70434 A1

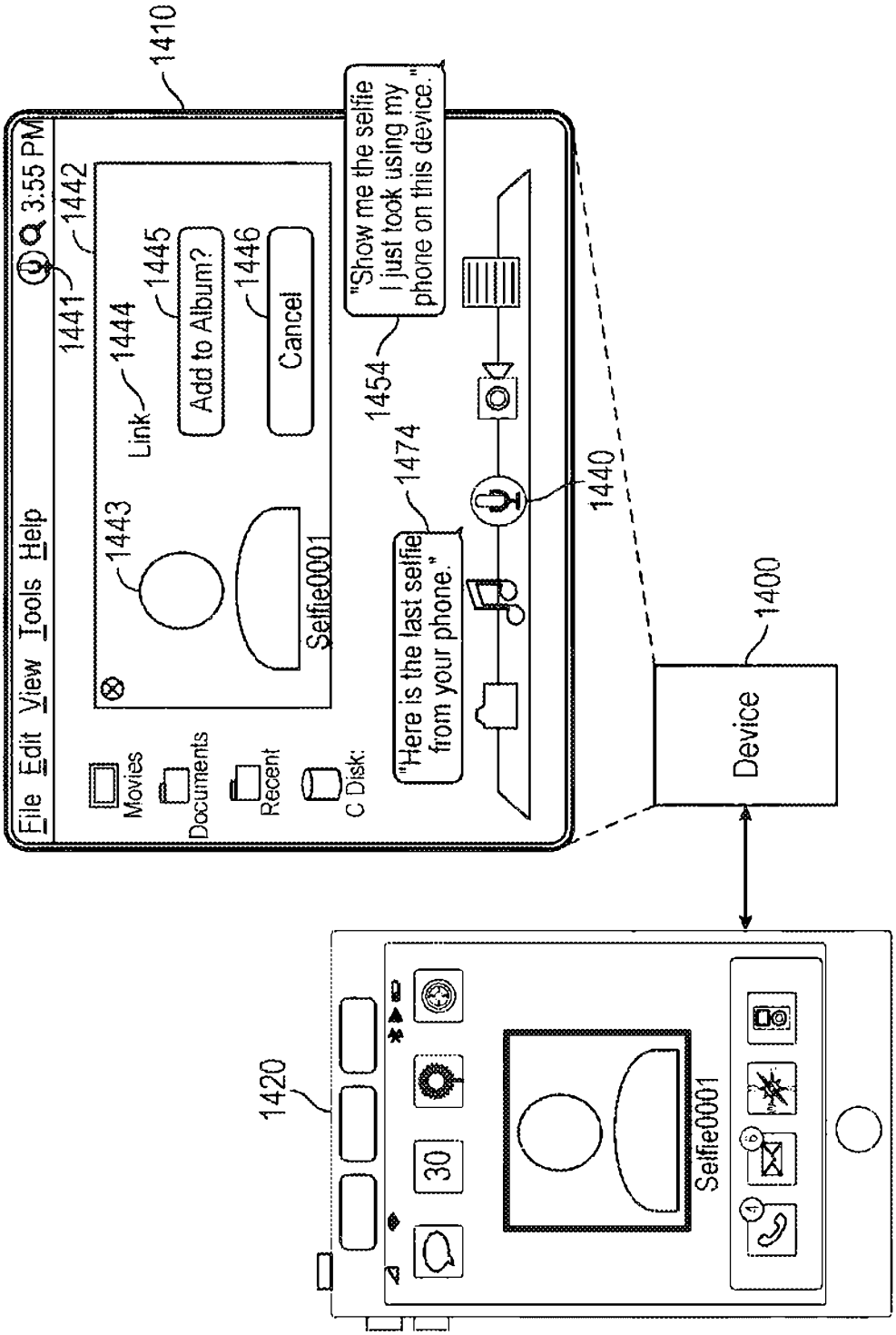


FIG. 14B

DK 2017 70434 A1

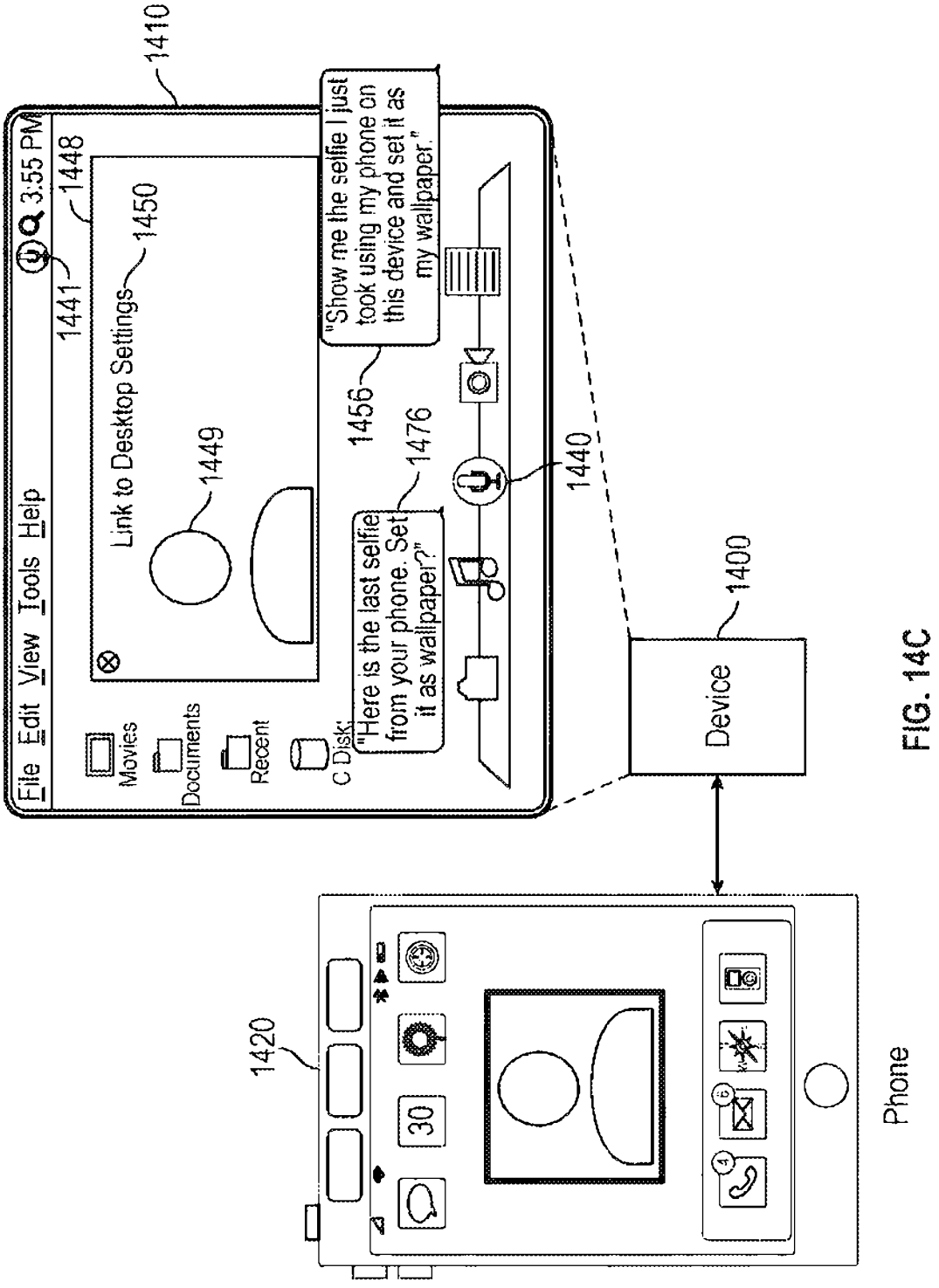


FIG. 14C

DK 2017 70434 A1

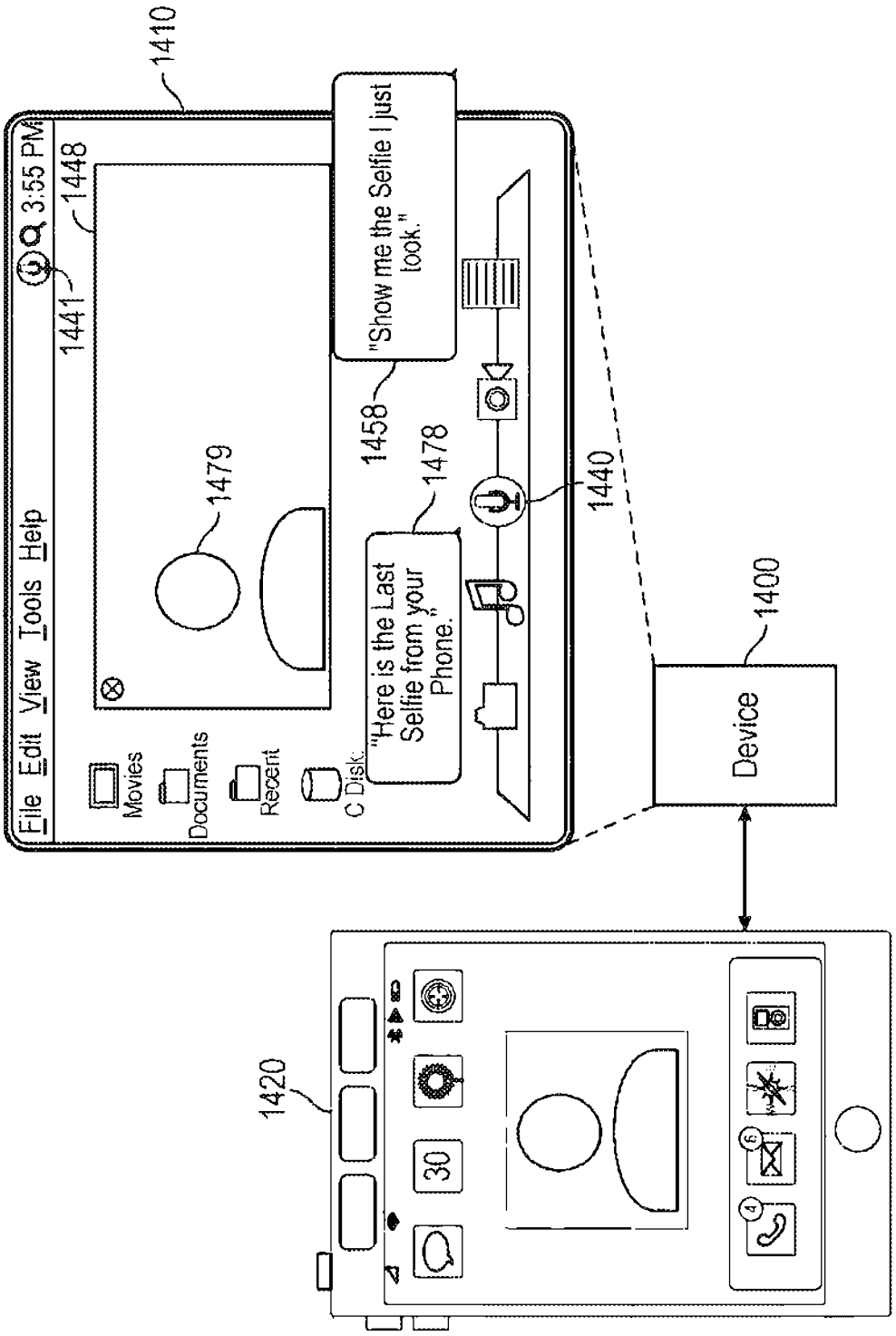


FIG. 14D

DK 2017 70434 A1

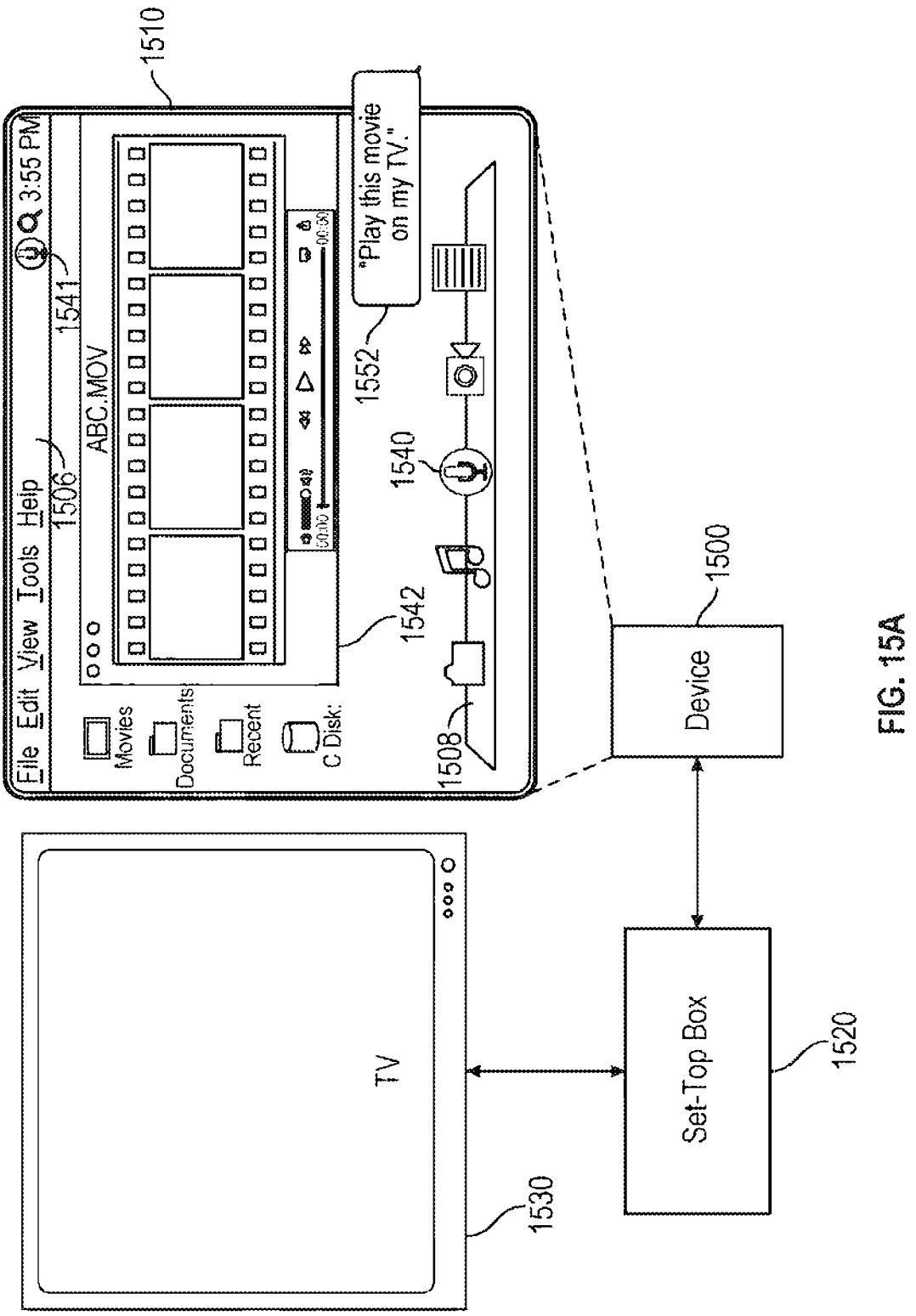


FIG. 15A

DK 2017 70434 A1

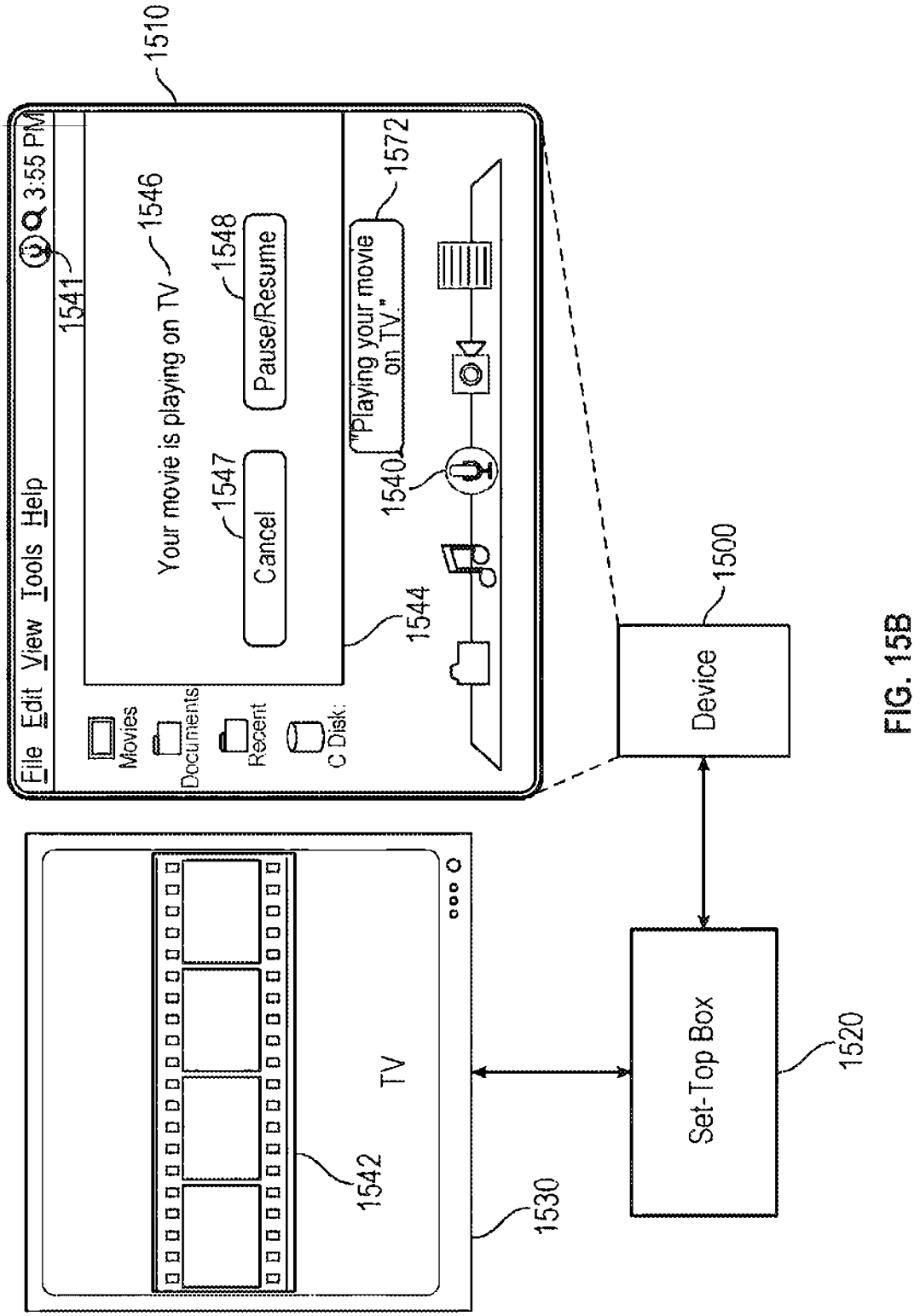


FIG. 15B

DK 2017 70434 A1

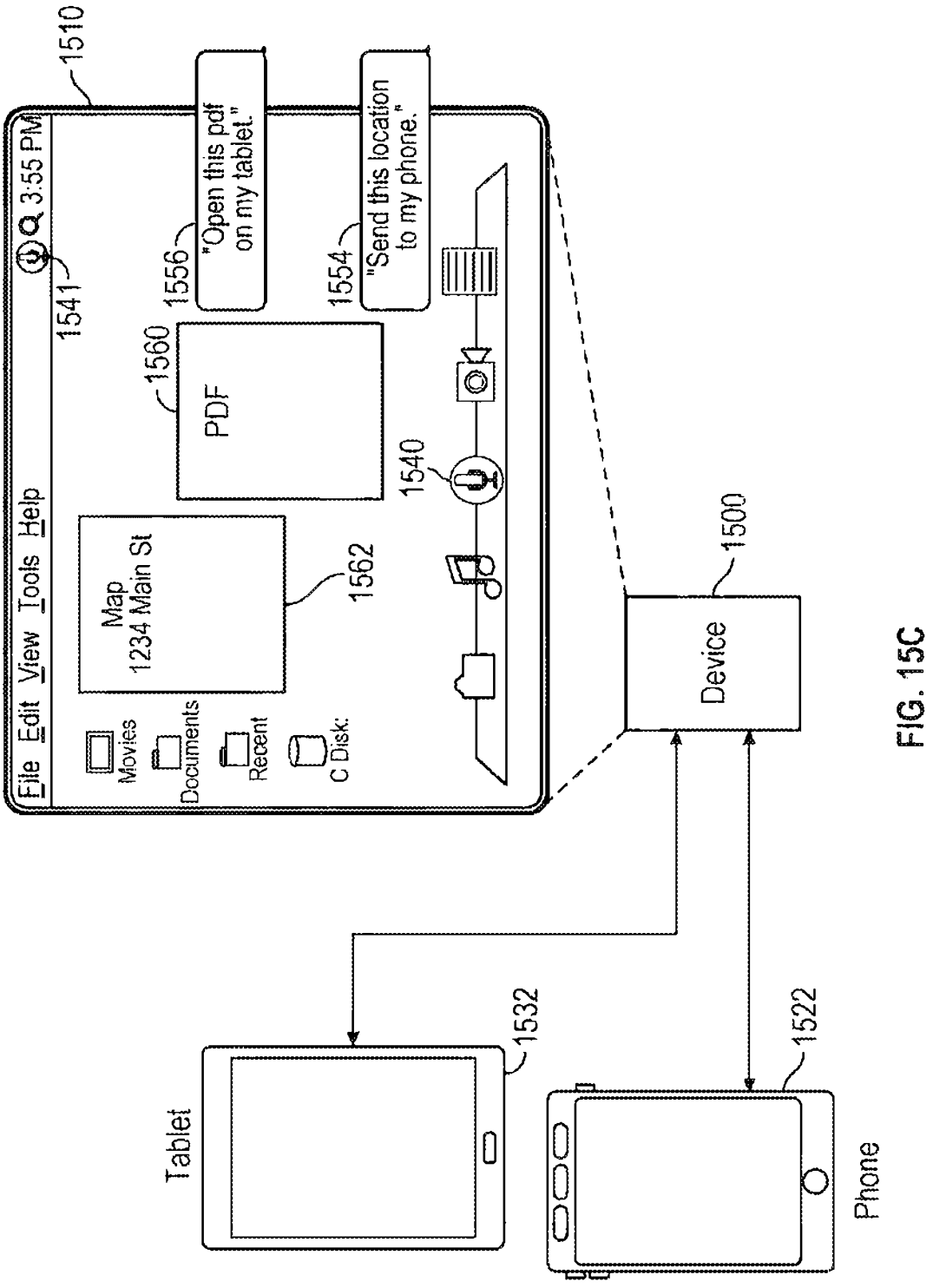


FIG. 15C

DK 2017 70434 A1

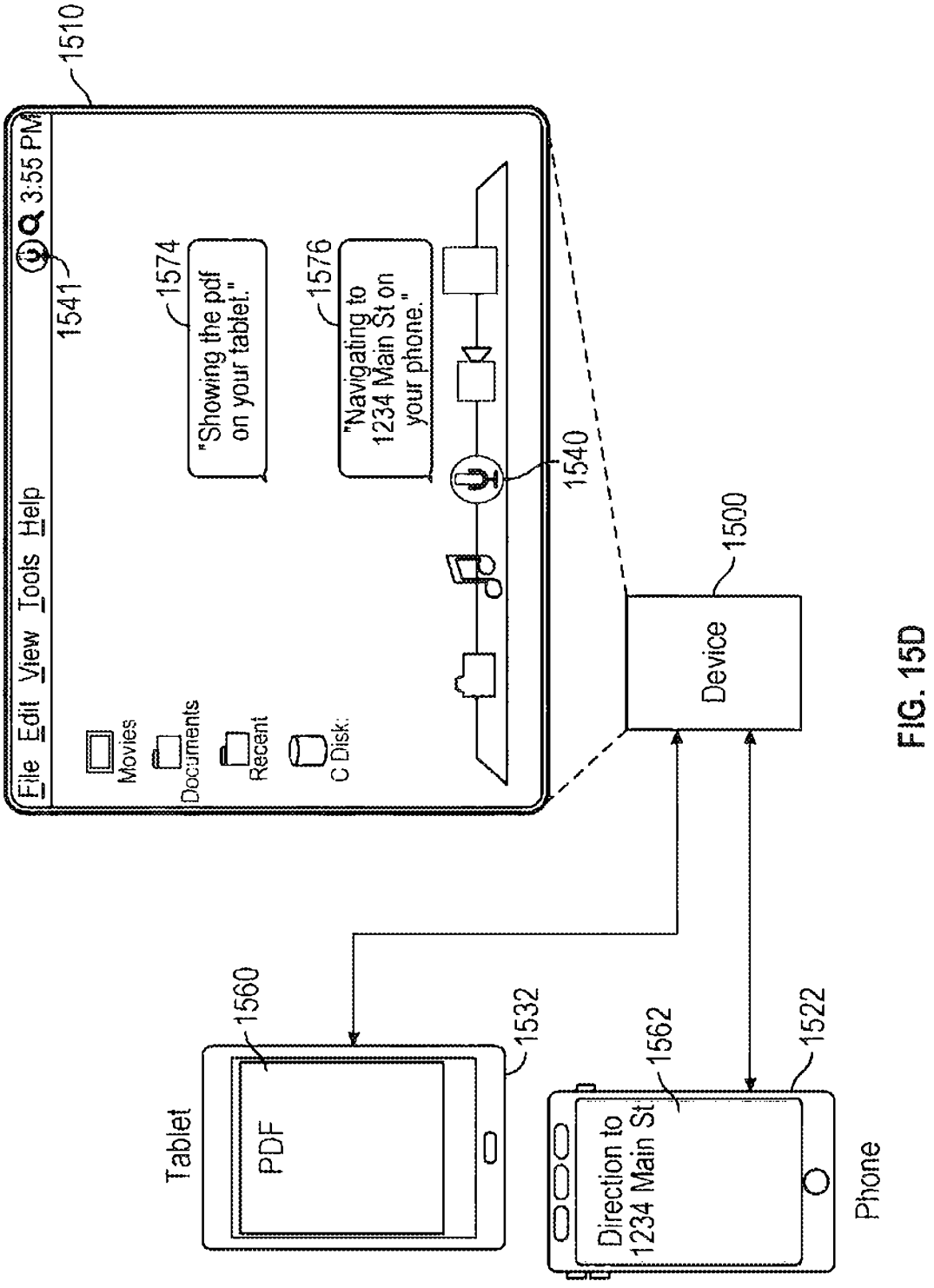


FIG. 15D

DK 2017 70434 A1

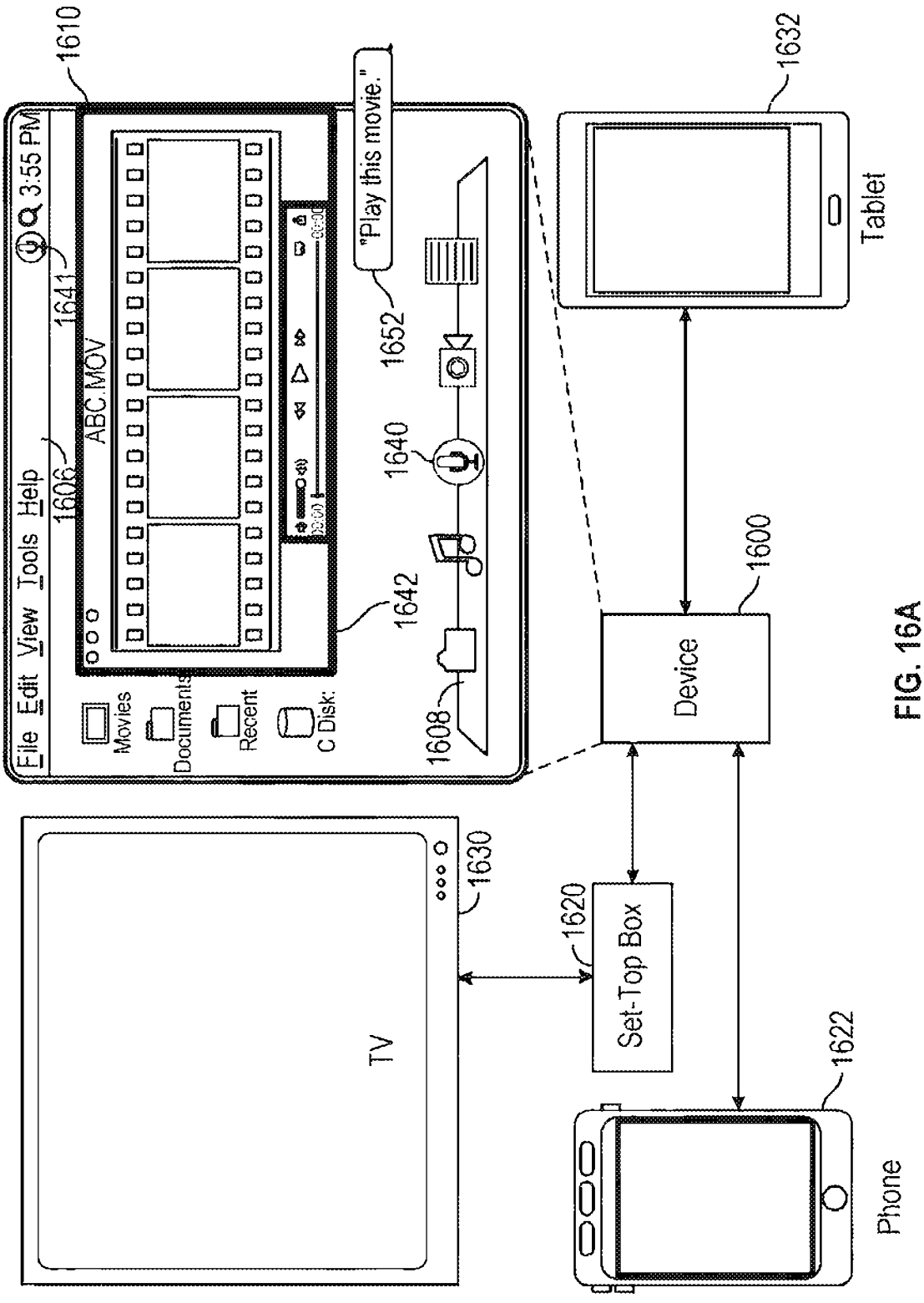


FIG. 16A

DK 2017 70434 A1

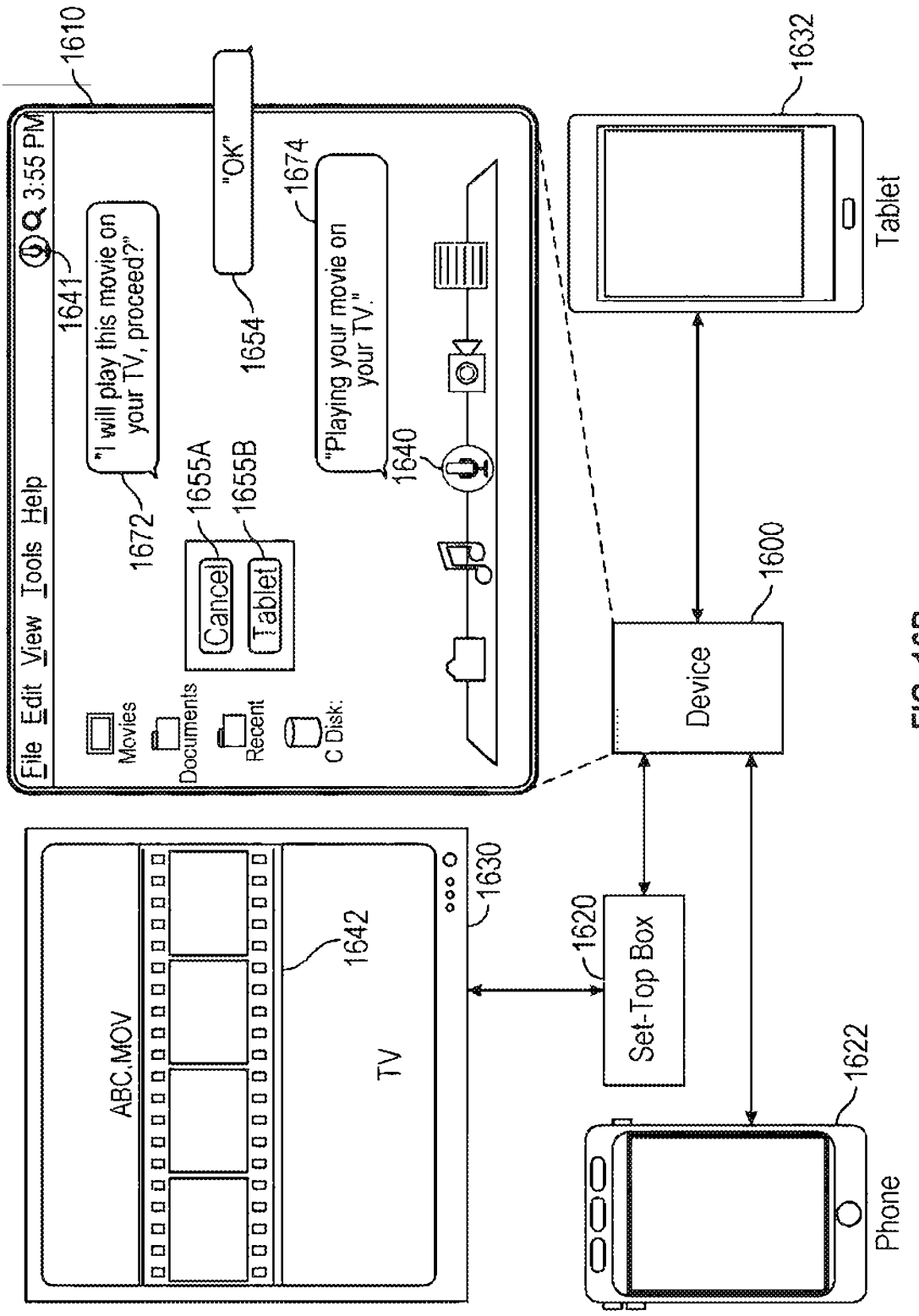


FIG. 16B

DK 2017 70434 A1

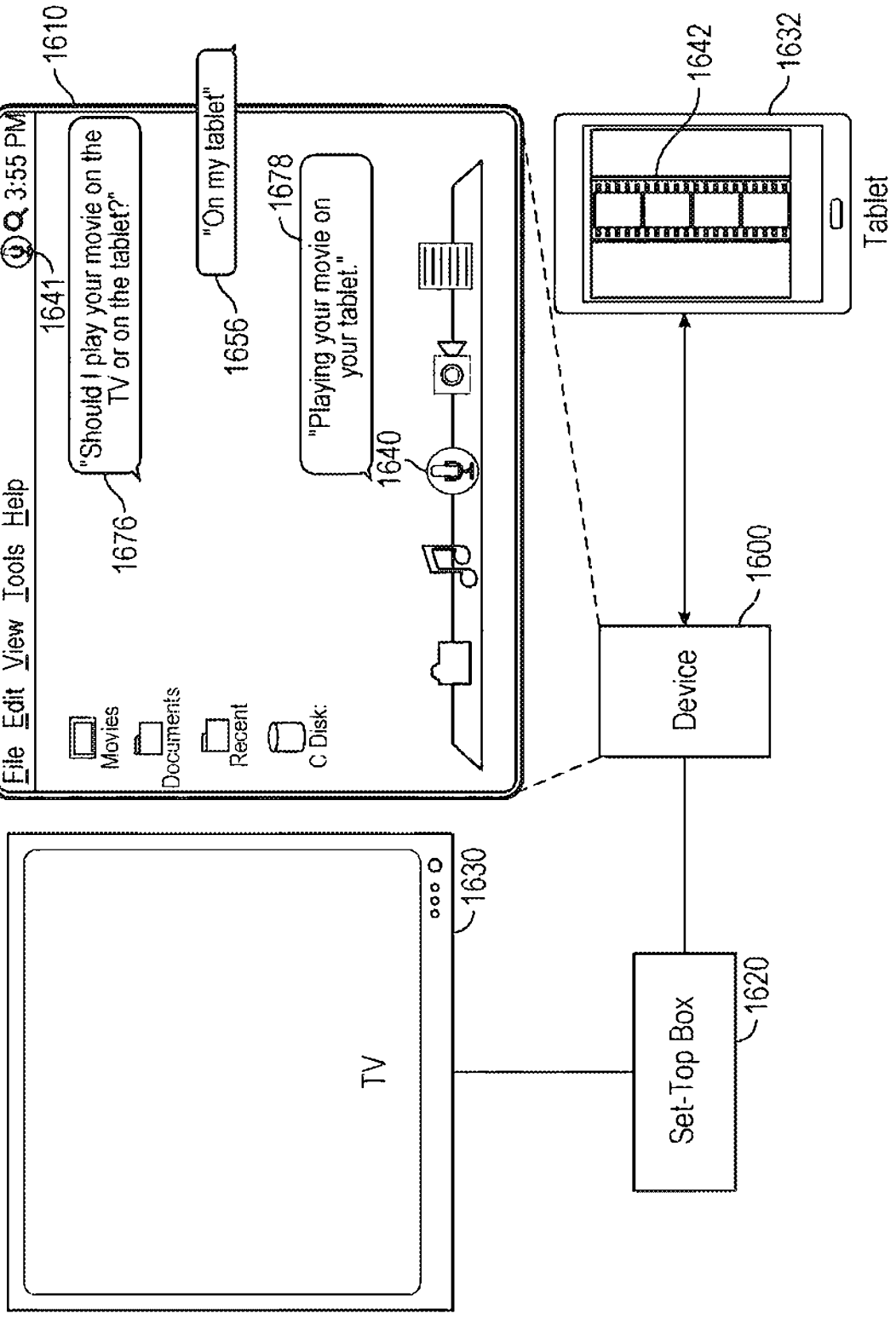


FIG. 16C

DK 2017 70434 A1

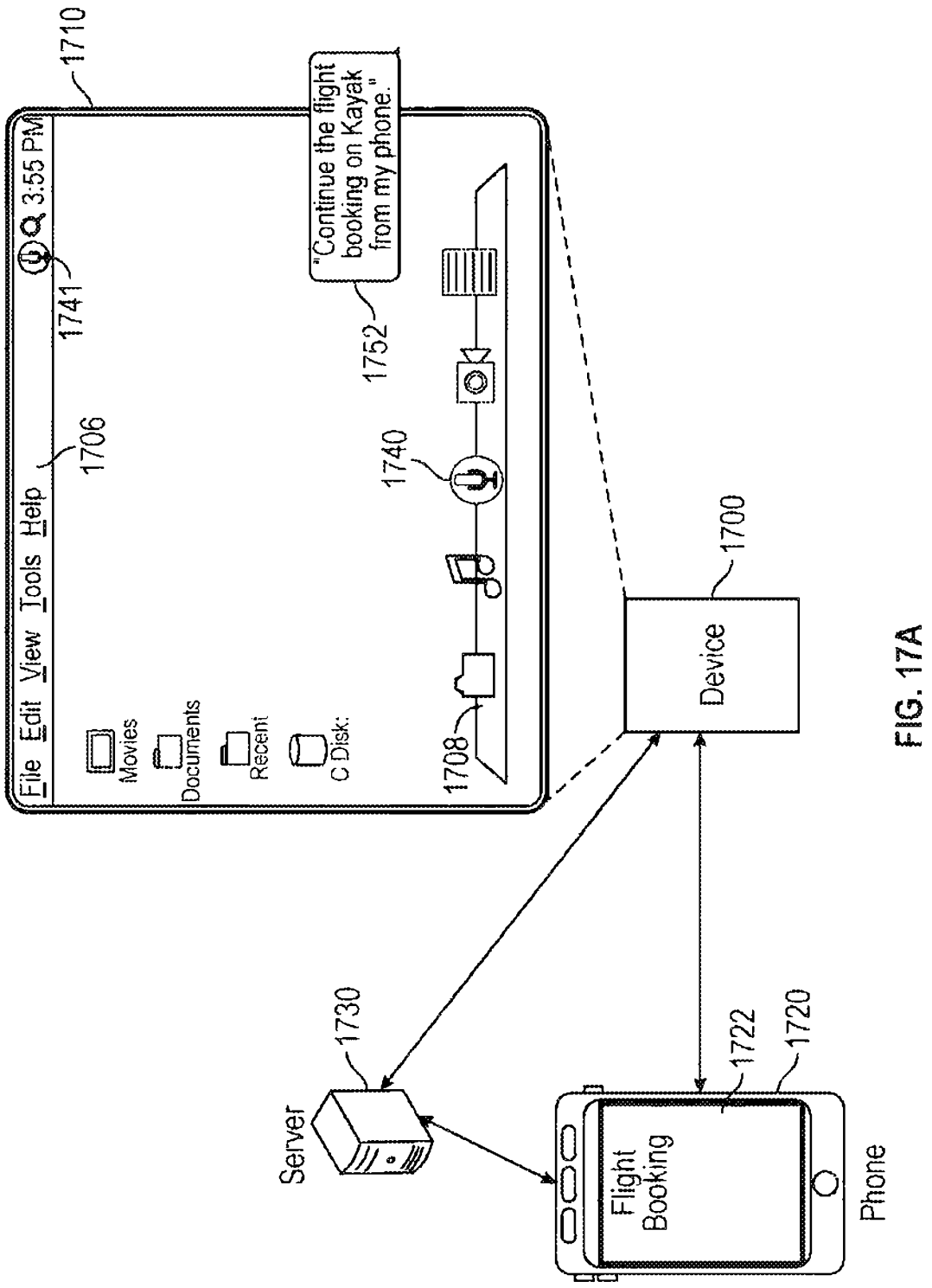


FIG. 17A

DK 2017 70434 A1

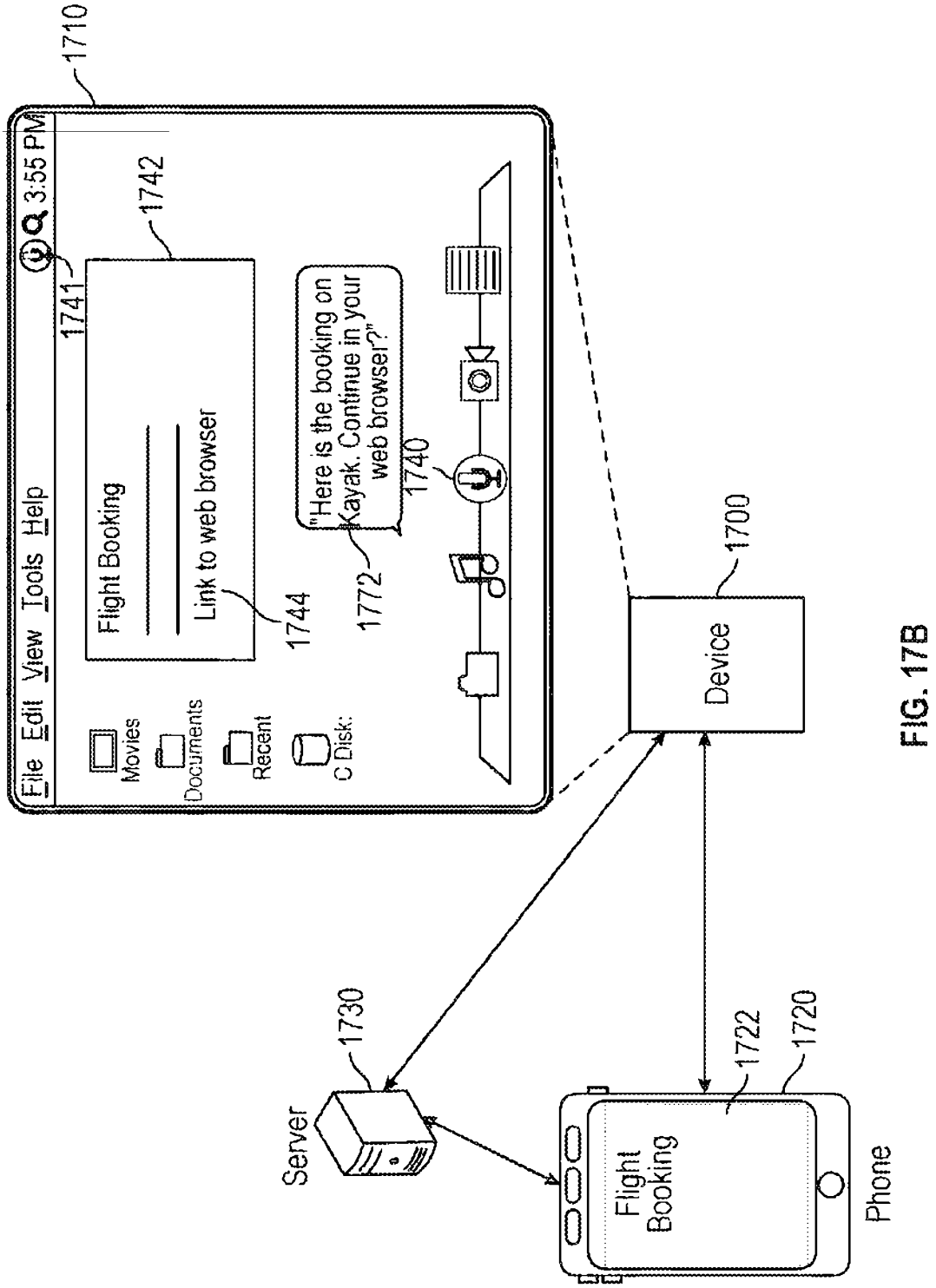


FIG. 17B

DK 2017 70434 A1

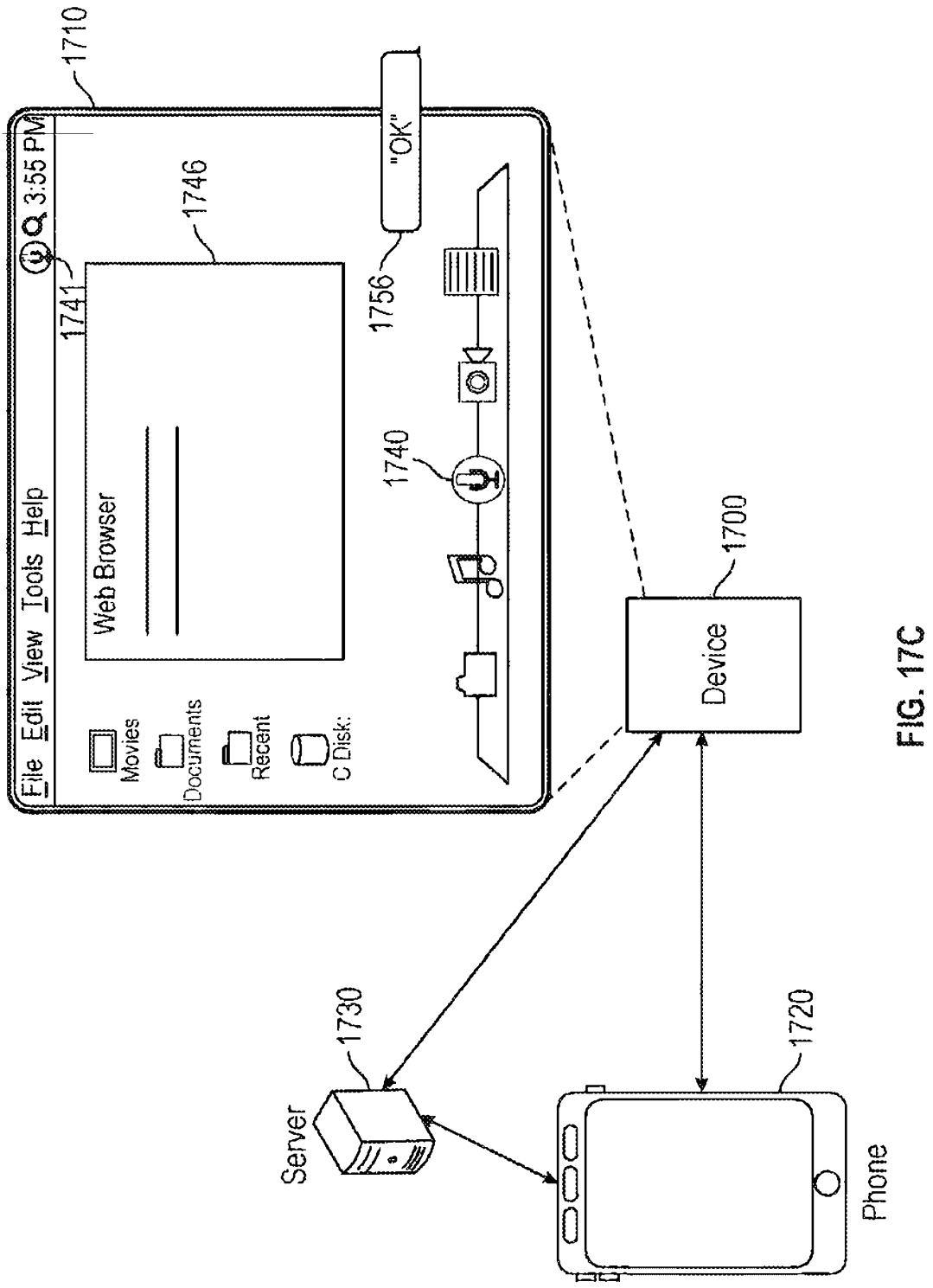


FIG. 17C

DK 2017 70434 A1

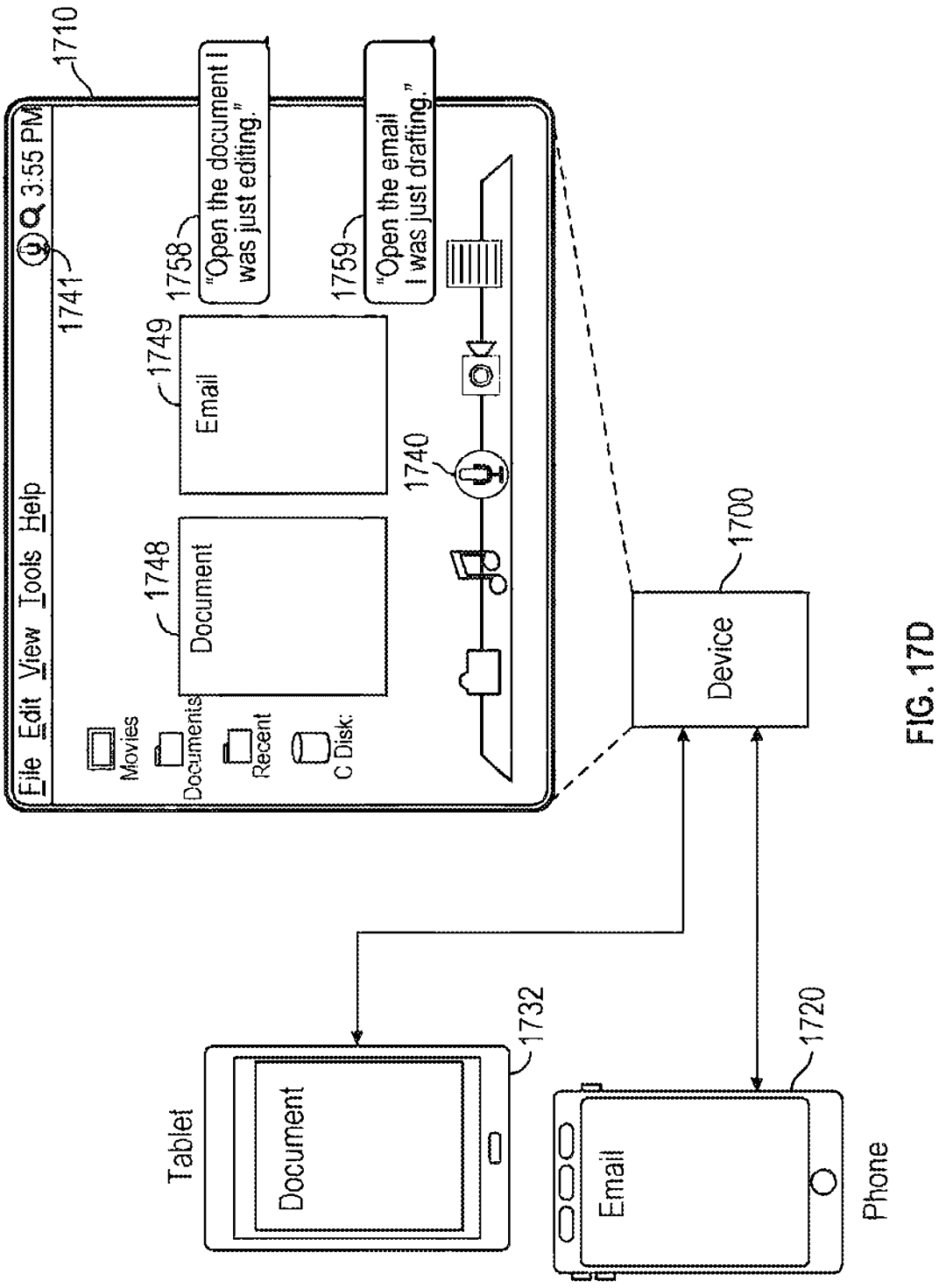


FIG. 17D

DK 2017 70434 A1

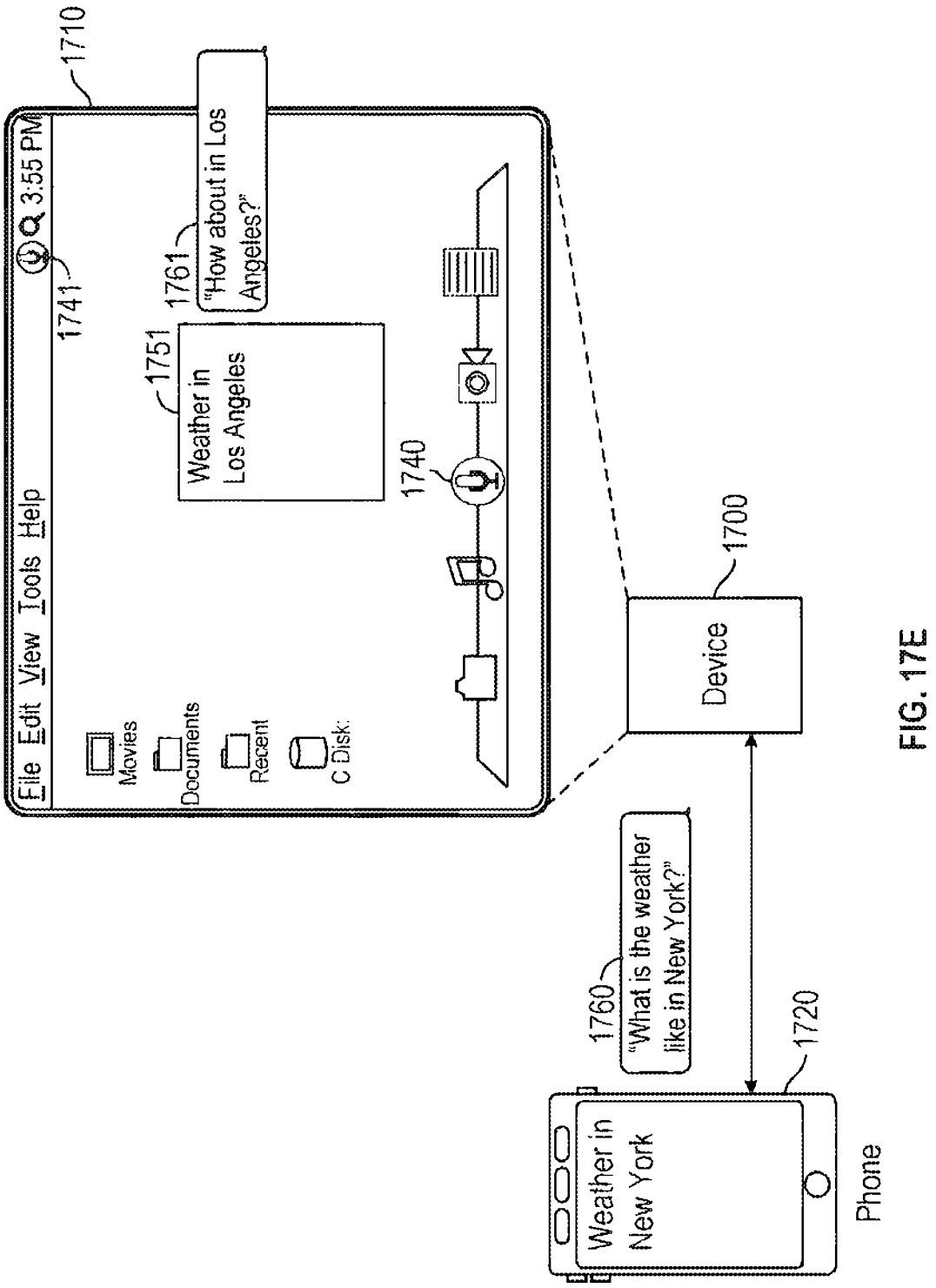


FIG. 17E

DK 2017 70434 A1

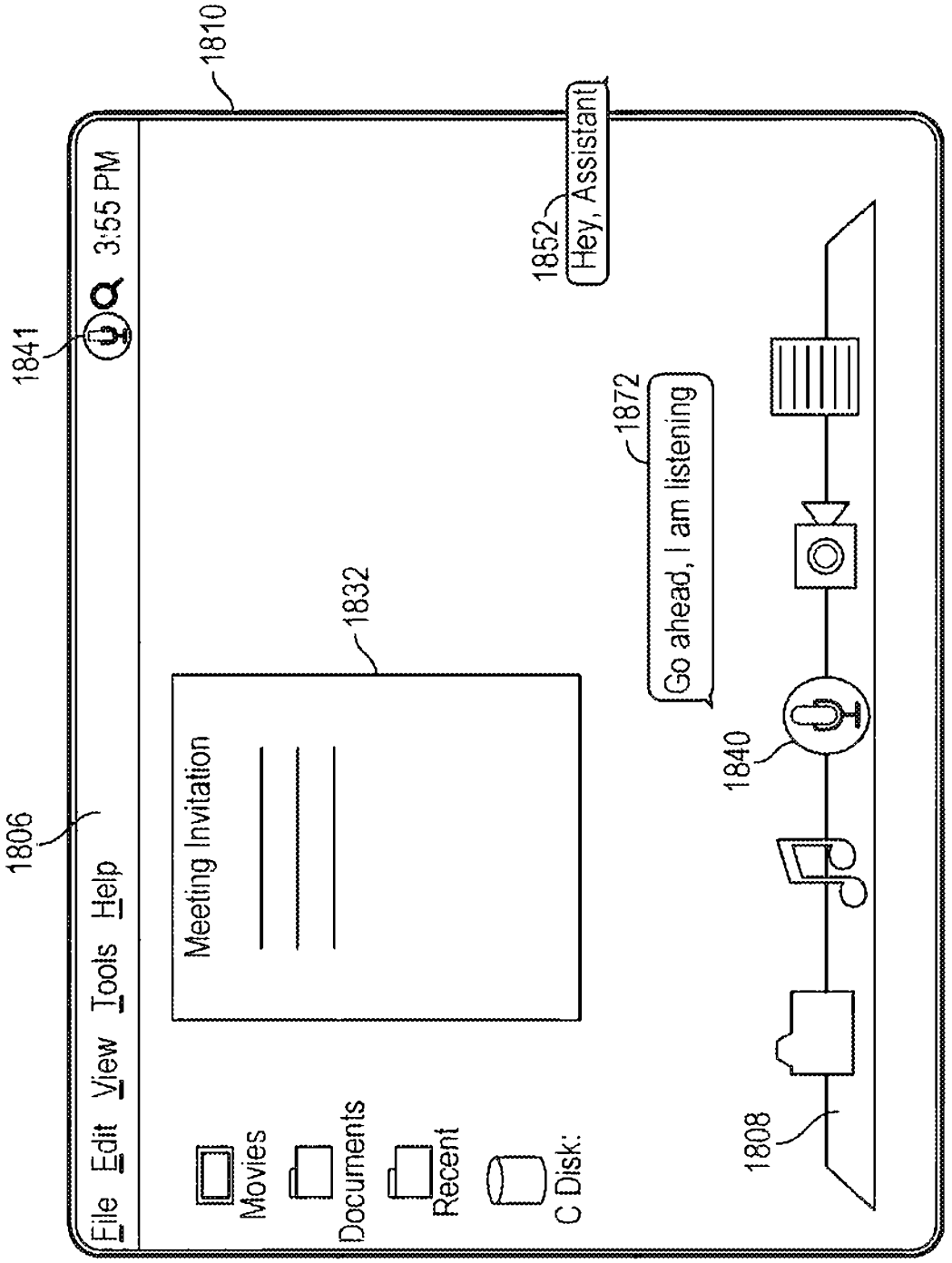


FIG. 18A

DK 2017 70434 A1

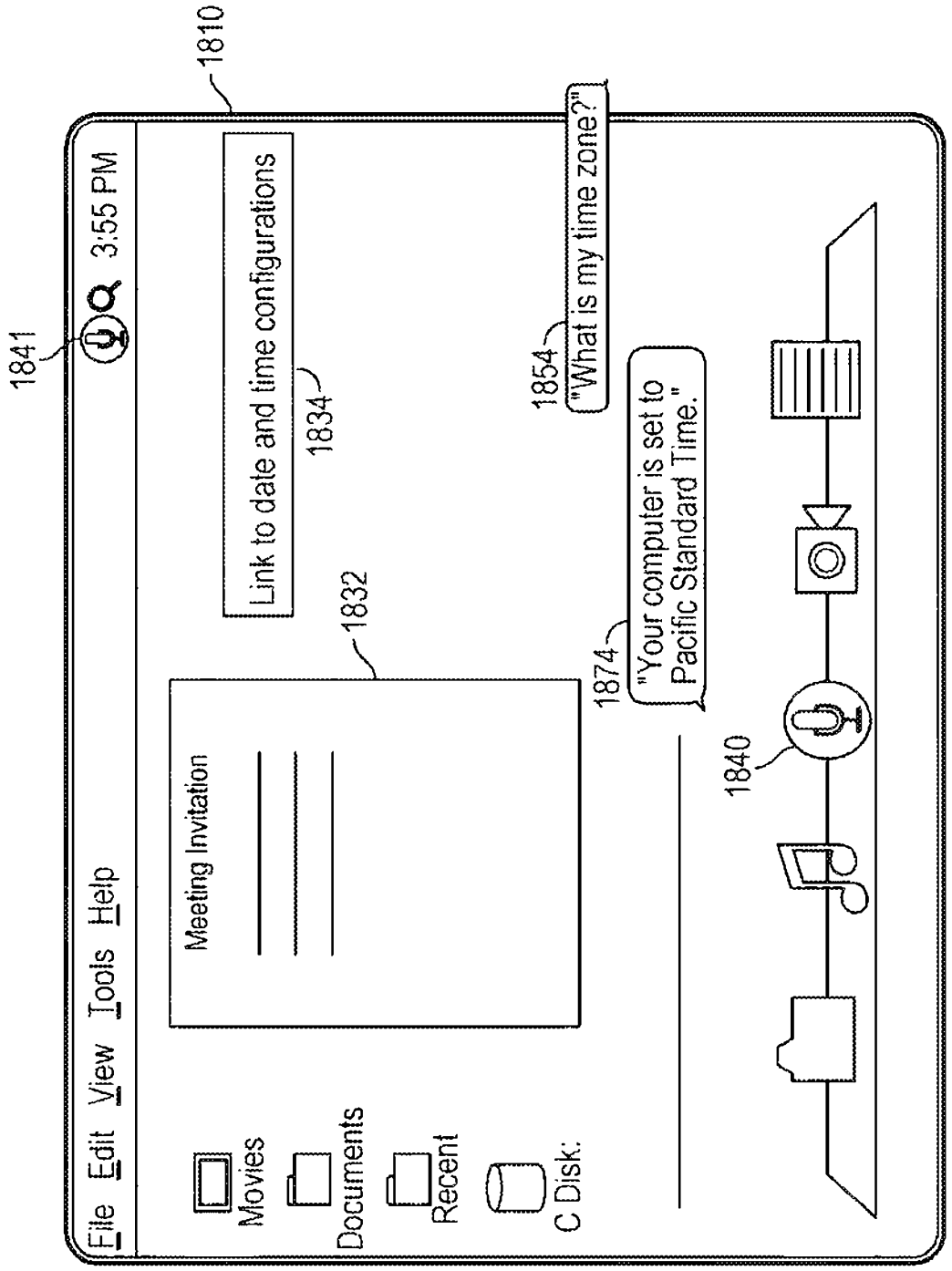


FIG. 18B

DK 2017 70434 A1

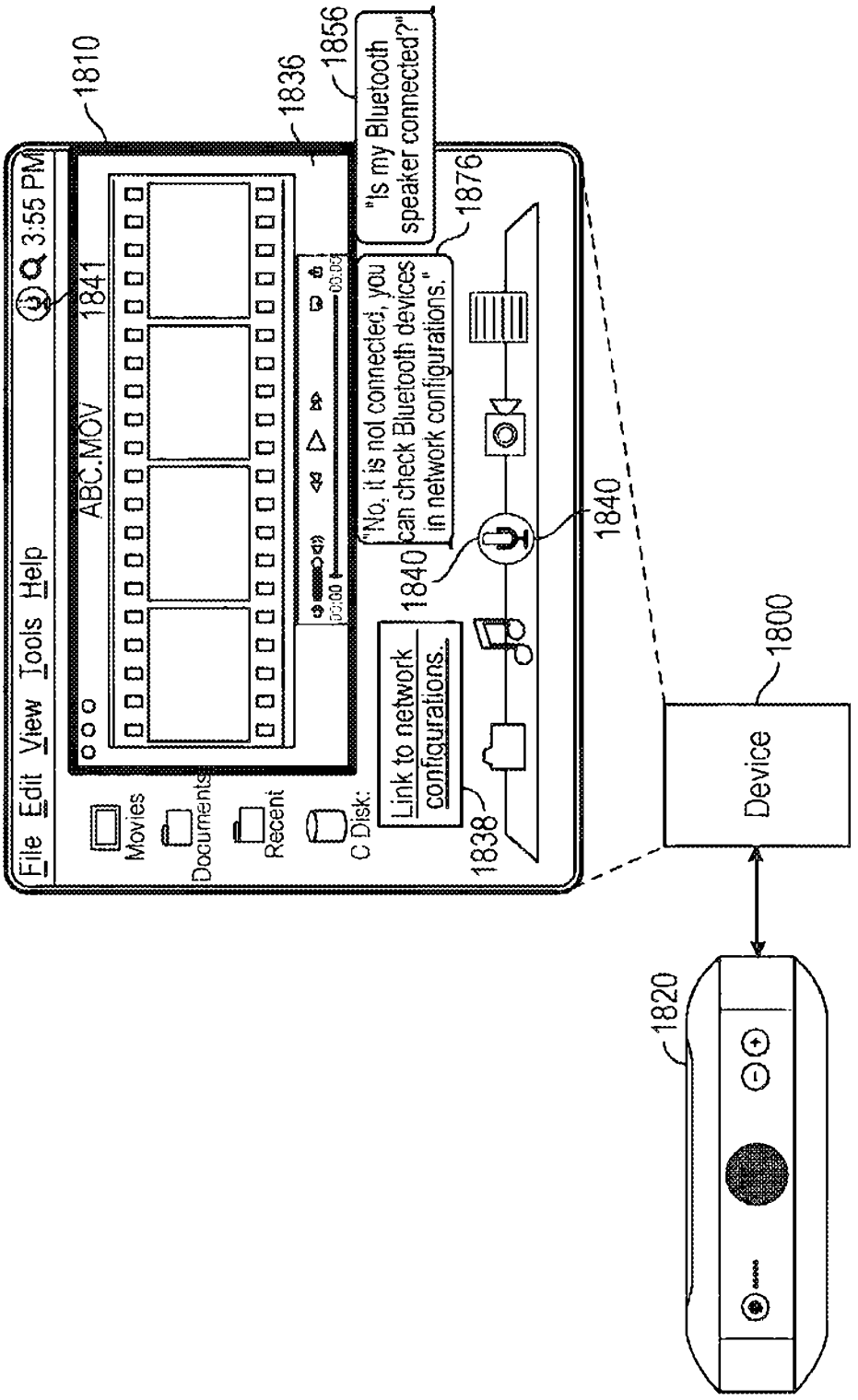


FIG. 18C

DK 2017 70434 A1

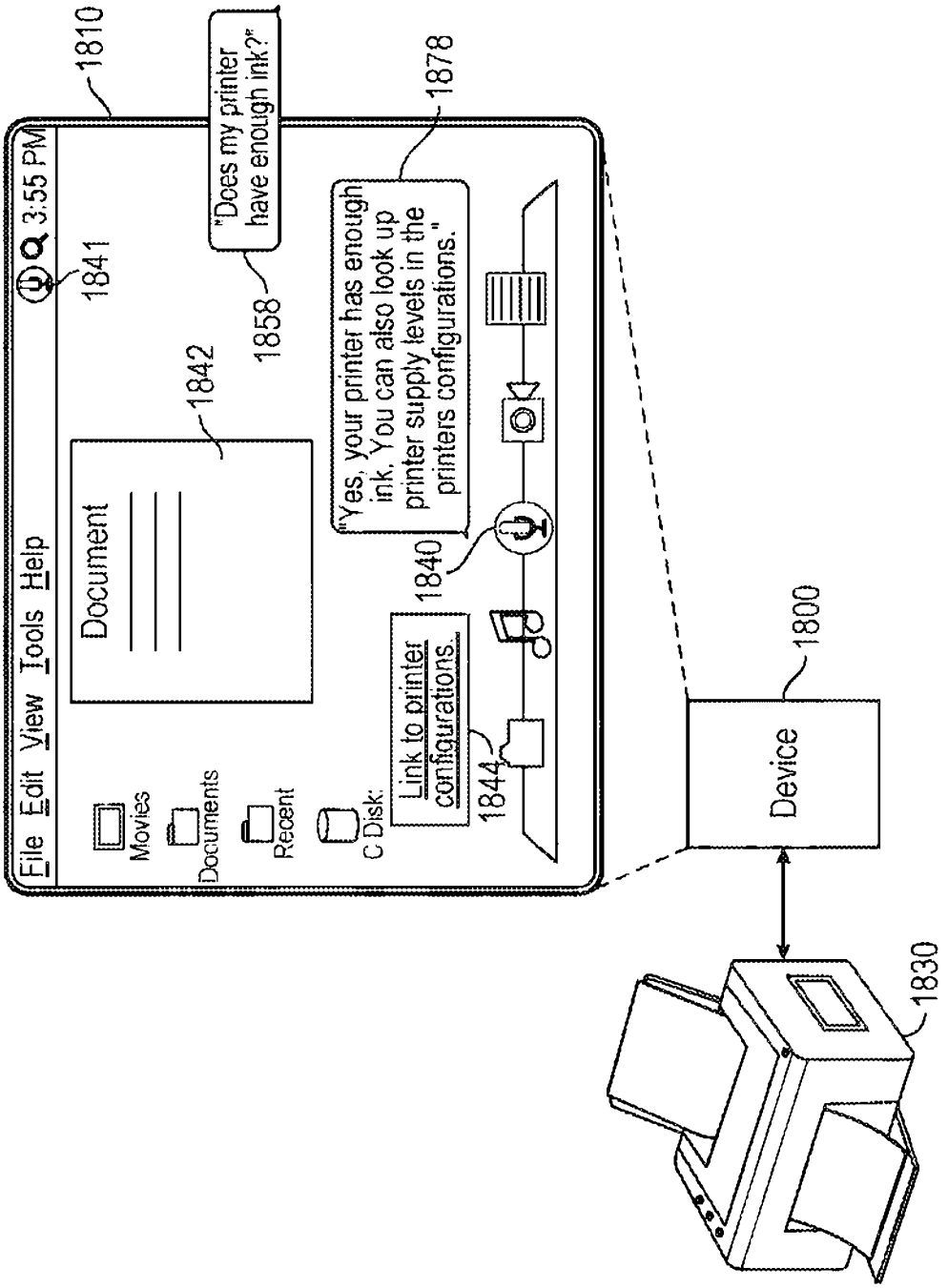


FIG. 18D

DK 2017 70434 A1

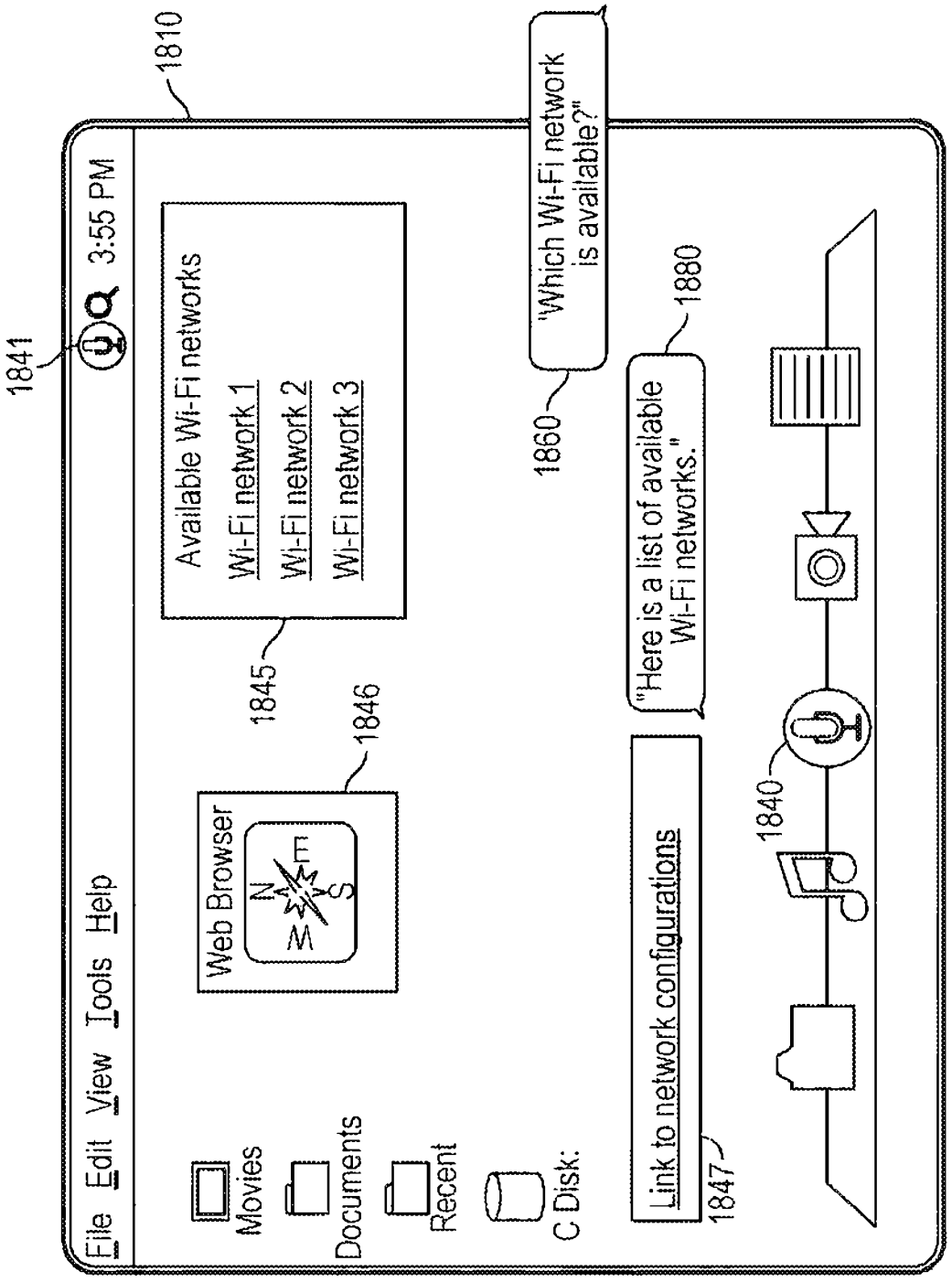


FIG. 18E

DK 2017 70434 A1

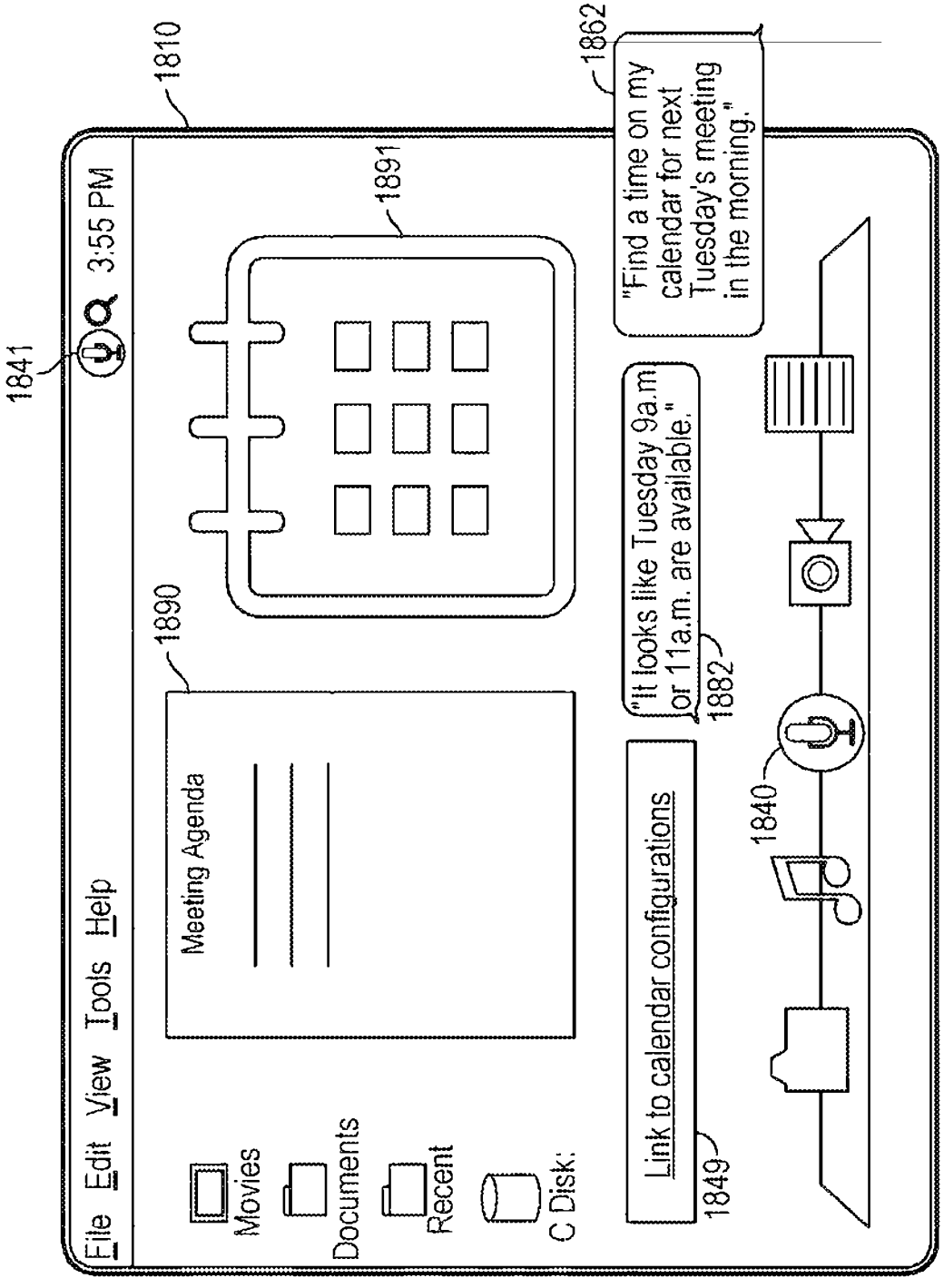


FIG. 18F

DK 2017 70434 A1

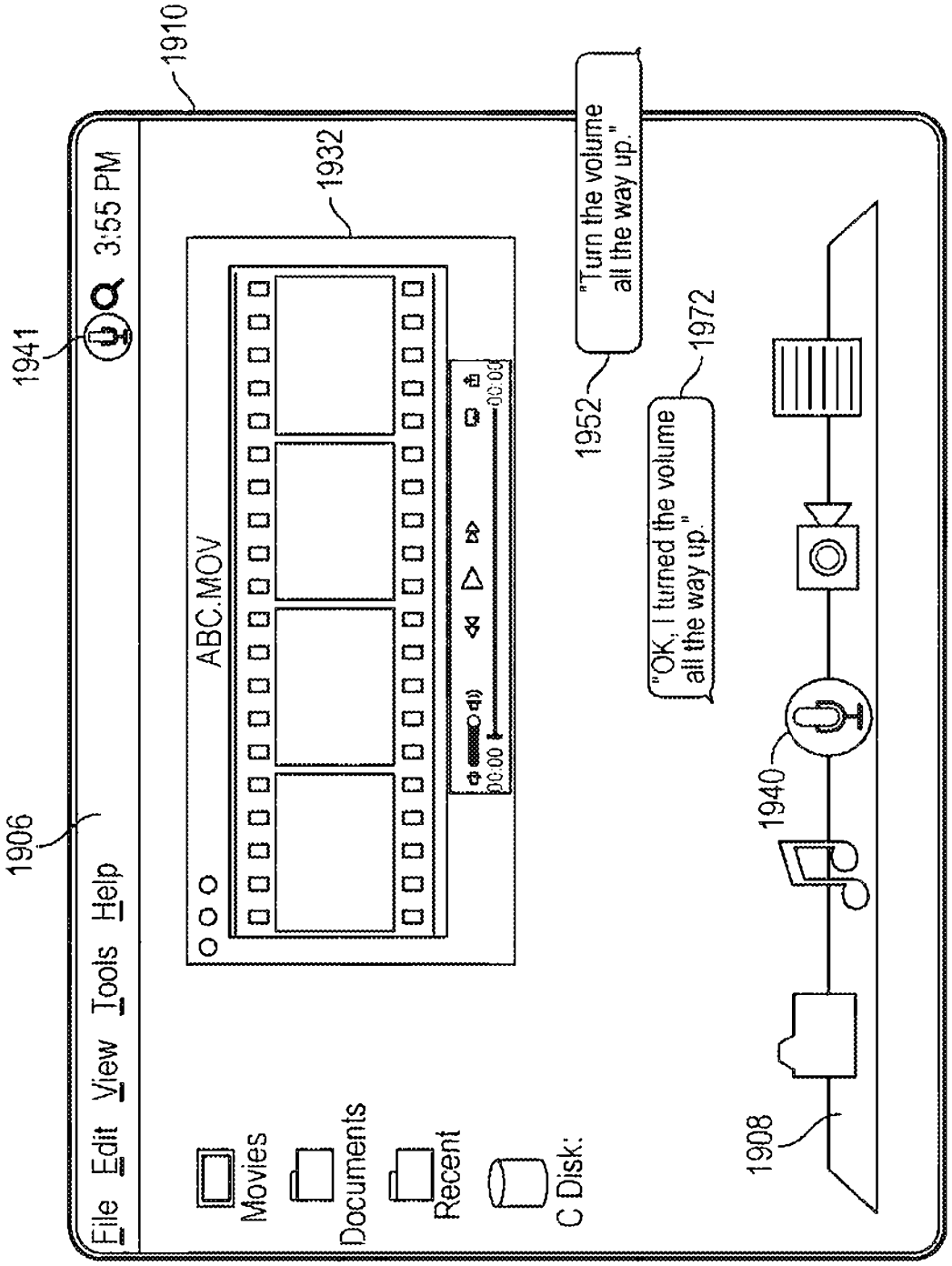


FIG. 19A

DK 2017 70434 A1

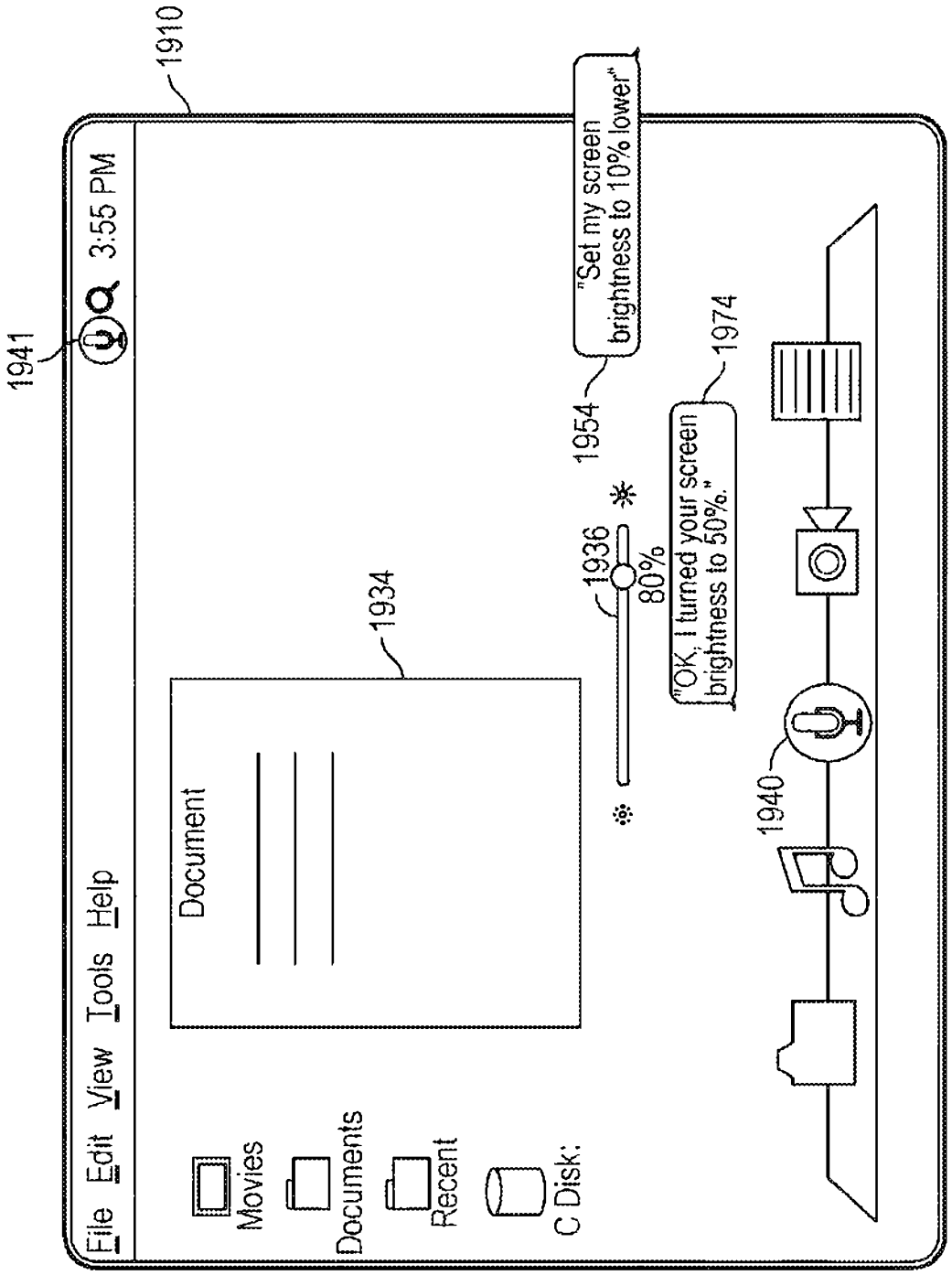


FIG. 19B

DK 2017 70434 A1

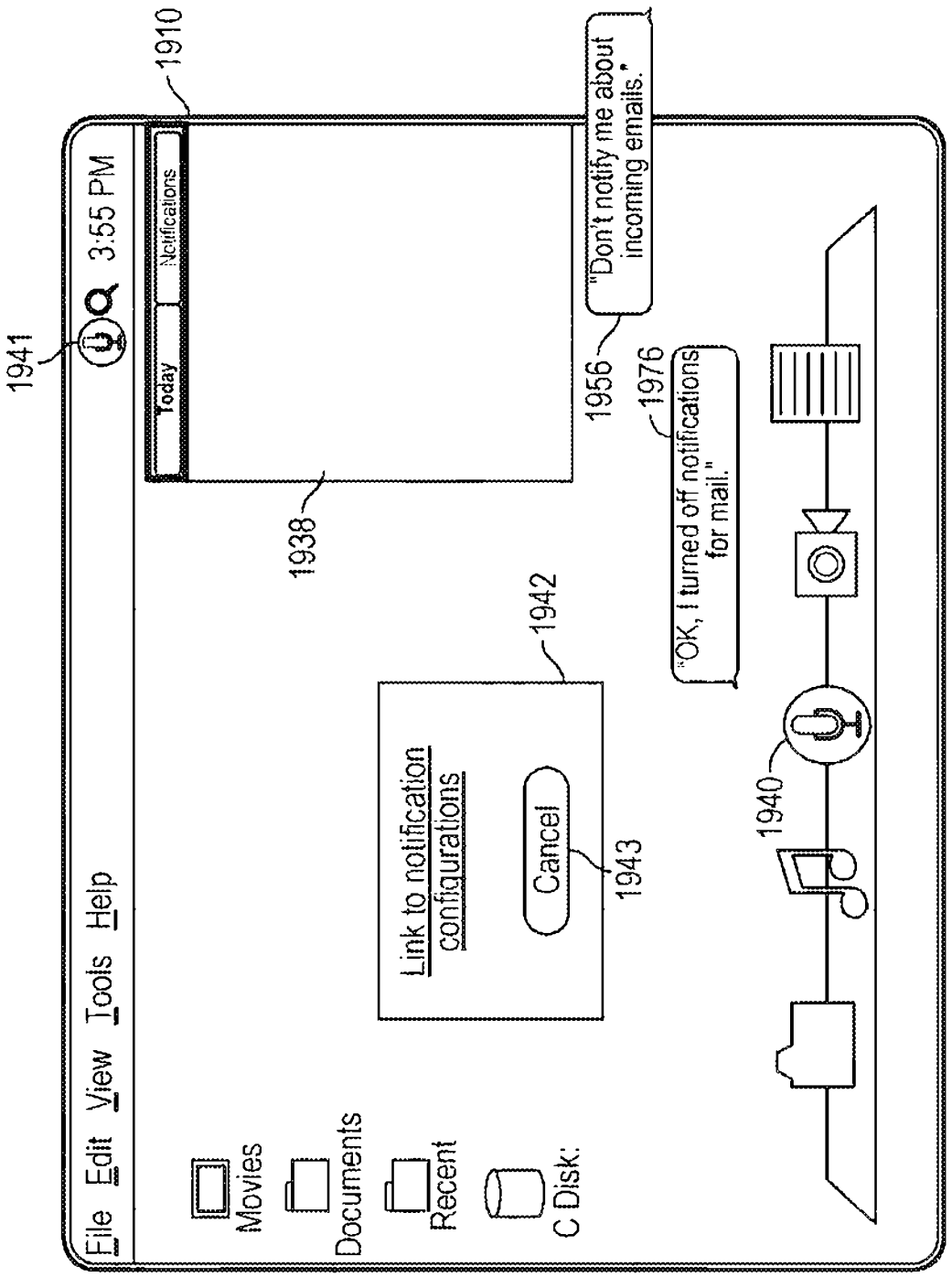


FIG. 19C

DK 2017 70434 A1

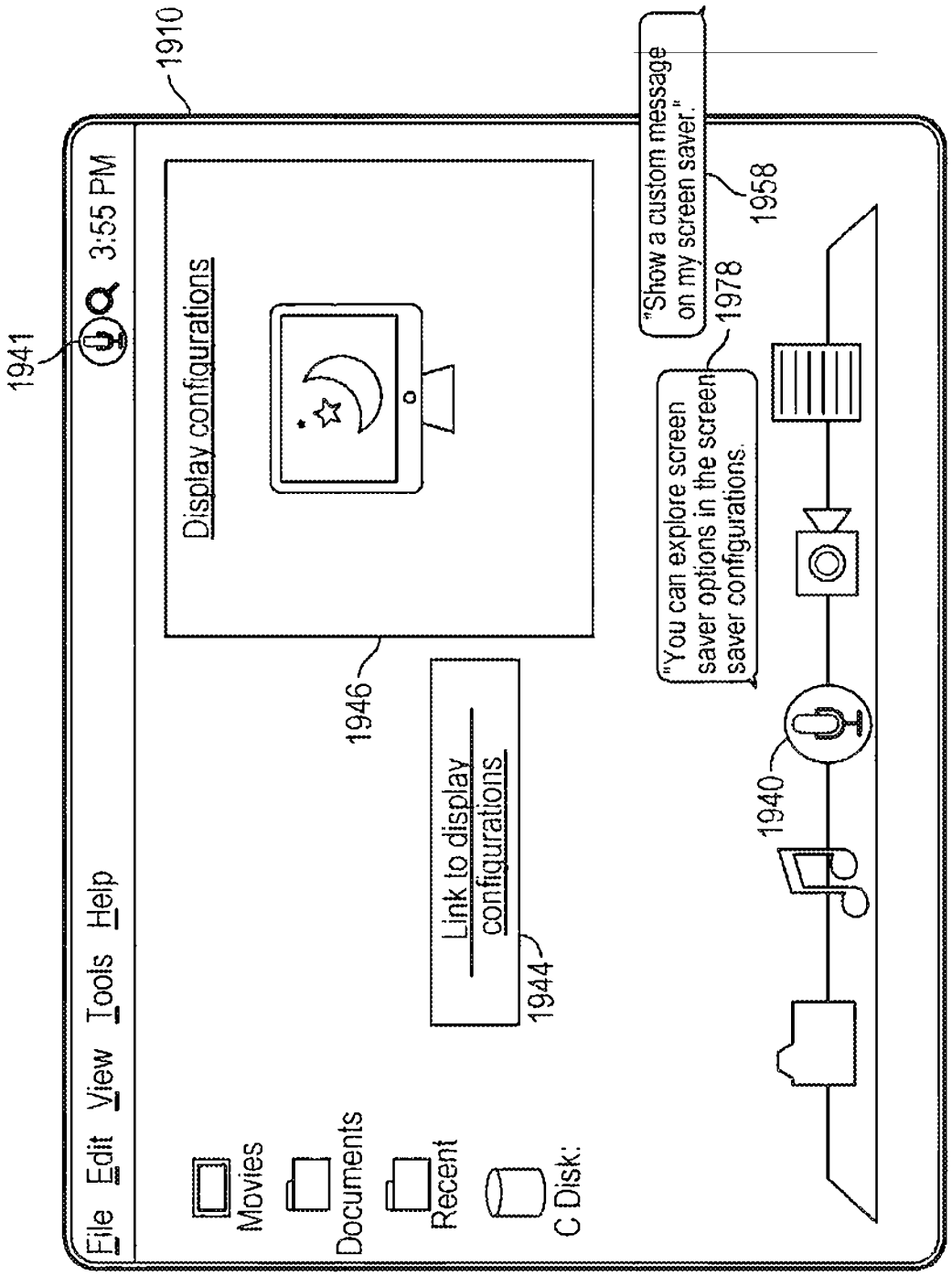


FIG. 19D

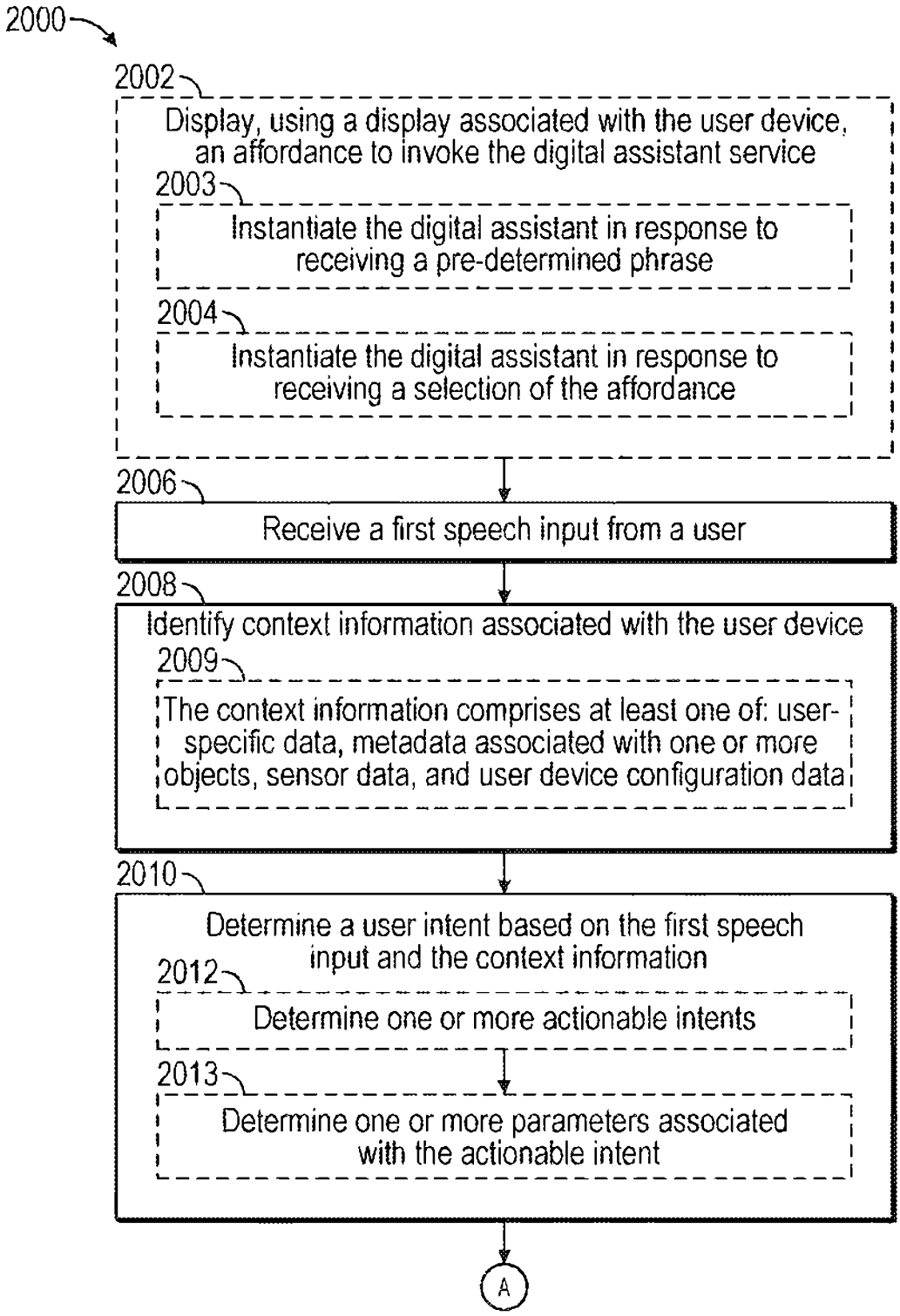

2000
2002
Display, using a display associated with the user device, an affordance to invoke the digital assistant service
2003
Instantiate the digital assistant in response to receiving a pre-determined phrase
2004
Instantiate the digital assistant in response to receiving a selection of the affordance
2006
Receive a first speech input from a user
2008
Identify context information associated with the user device
2009
The context information comprises at least one of: user-specific data, metadata associated with one or more objects, sensor data, and user device configuration data
2010
Determine a user intent based on the first speech input and the context information
2012
Determine one or more actionable intents
2013
Determine one or more parameters associated with the actionable intent
A

FIG. 20A

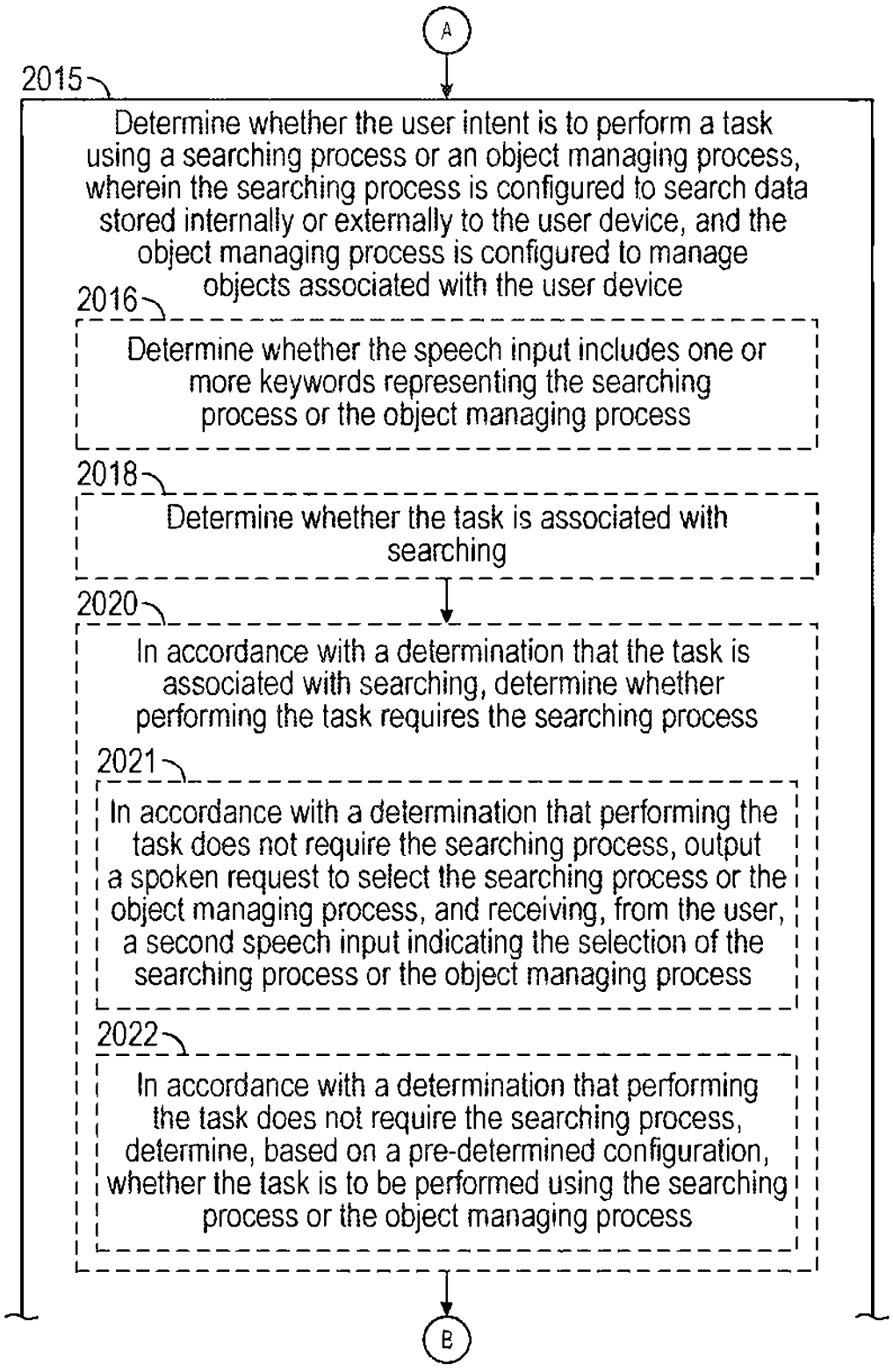

A
2015
Determine whether the user intent is to perform a task using a searching process or an object managing process, wherein the searching process is configured to search data stored internally or externally to the user device, and the object managing process is configured to manage objects associated with the user device
2016
Determine whether the speech input includes one or more keywords representing the searching process or the object managing process
2018
Determine whether the task is associated with searching
2020
In accordance with a determination that the task is associated with searching, determine whether performing the task requires the searching process
2021
In accordance with a determination that performing the task does not require the searching process, output a spoken request to select the searching process or the object managing process, and receiving, from the user, a second speech input indicating the selection of the searching process or the object managing process
2022
In accordance with a determination that performing the task does not require the searching process, determine, based on a pre-determined configuration, whether the task is to be performed using the searching process or the object managing process
B

FIG. 20B

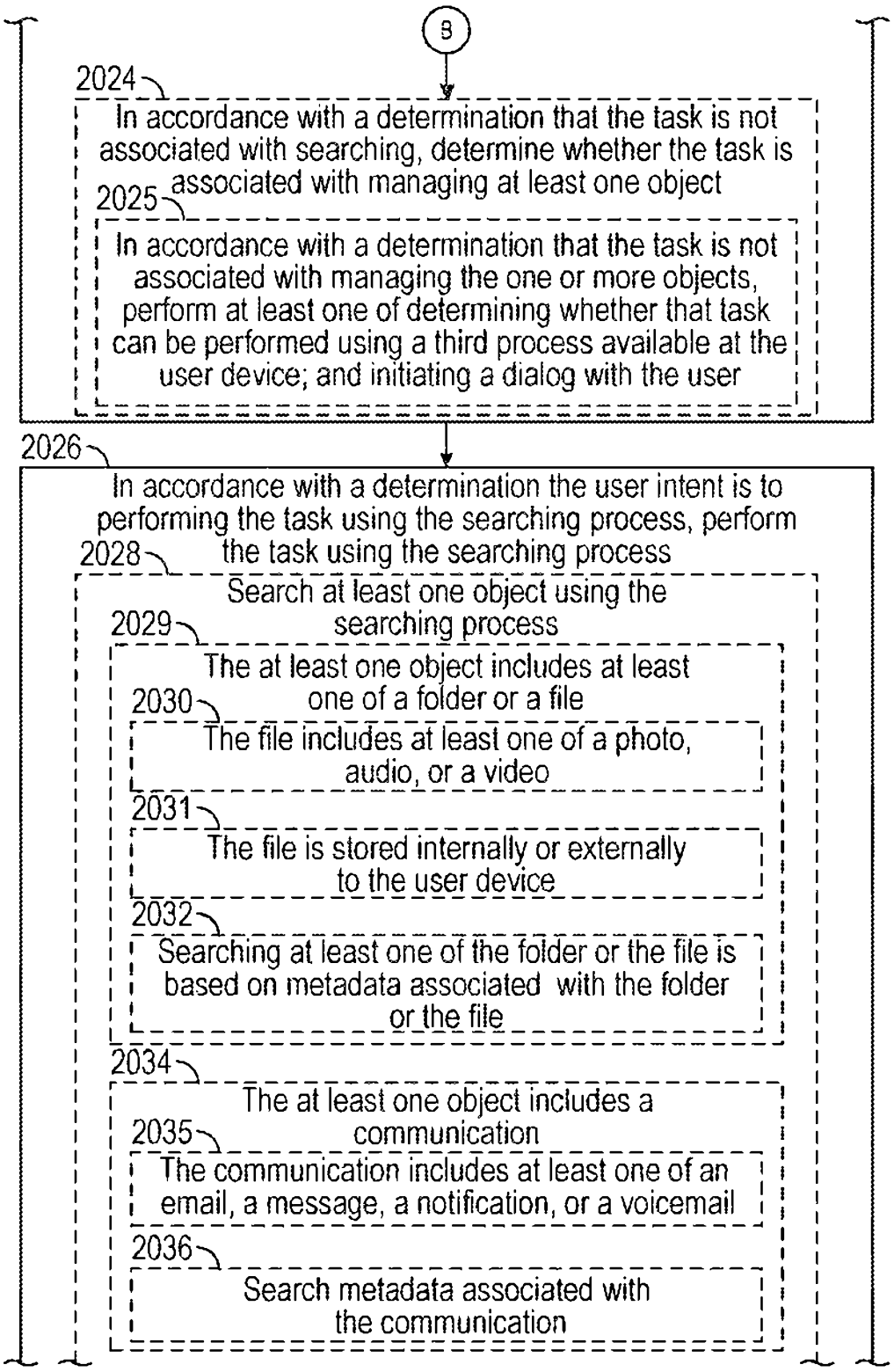

B
2024
In accordance with a determination that the task is not associated with searching, determine whether the task is associated with managing at least one object
2025
In accordance with a determination that the task is not associated with managing the one or more objects, perform at least one of determining whether that task can be performed using a third process available at the user device; and initiating a dialog with the user
2026
In accordance with a determination the user intent is to performing the task using the searching process, perform the task using the searching process
2028
Search at least one object using the searching process
2029
The at least one object includes at least one of a folder or a file
2030
The file includes at least one of a photo, audio, or a video
2031
The file is stored internally or externally to the user device
2032
Searching at least one of the folder or the file is based on metadata associated with the folder or the file
2034
The at least one object includes a communication
2035
The communication includes at least one of an email, a message, a notification, or a voicemail
2036
Search metadata associated with the communication

FIG. 20C

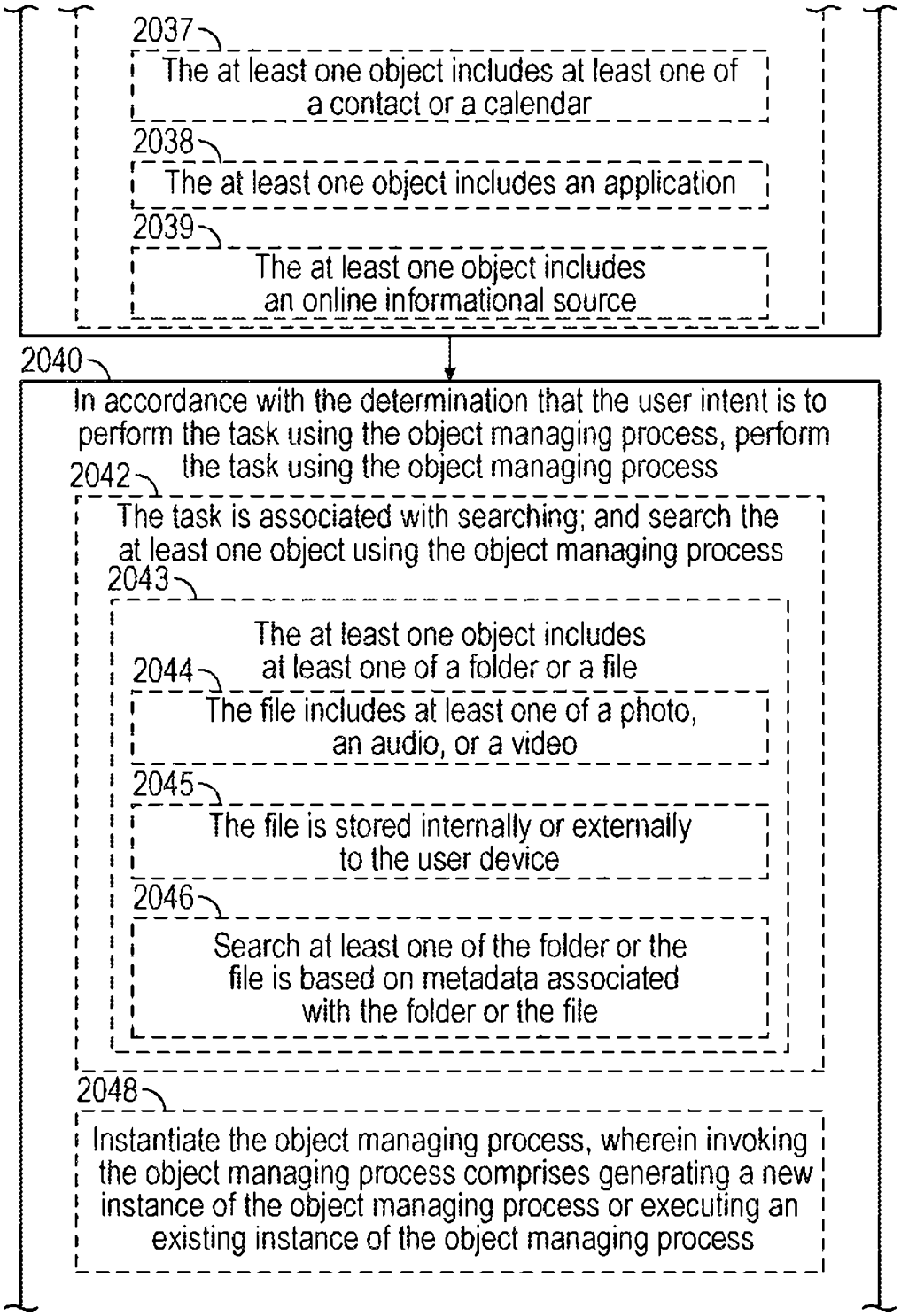

2037
The at least one object includes at least one of a contact or a calendar
2038
The at least one object includes an application
2039
The at least one object includes an online informational source
2040
In accordance with the determination that the user intent is to perform the task using the object managing process, perform the task using the object managing process
2042
The task is associated with searching; and search the at least one object using the object managing process
2043
The at least one object includes at least one of a folder or a file
2044
The file includes at least one of a photo, an audio, or a video
2045
The file is stored internally or externally to the user device
2046
Search at least one of the folder or the file is based on metadata associated with the folder or the file
2048
Instantiate the object managing process, wherein invoking the object managing process comprises generating a new instance of the object managing process or executing an existing instance of the object managing process

FIG. 20D

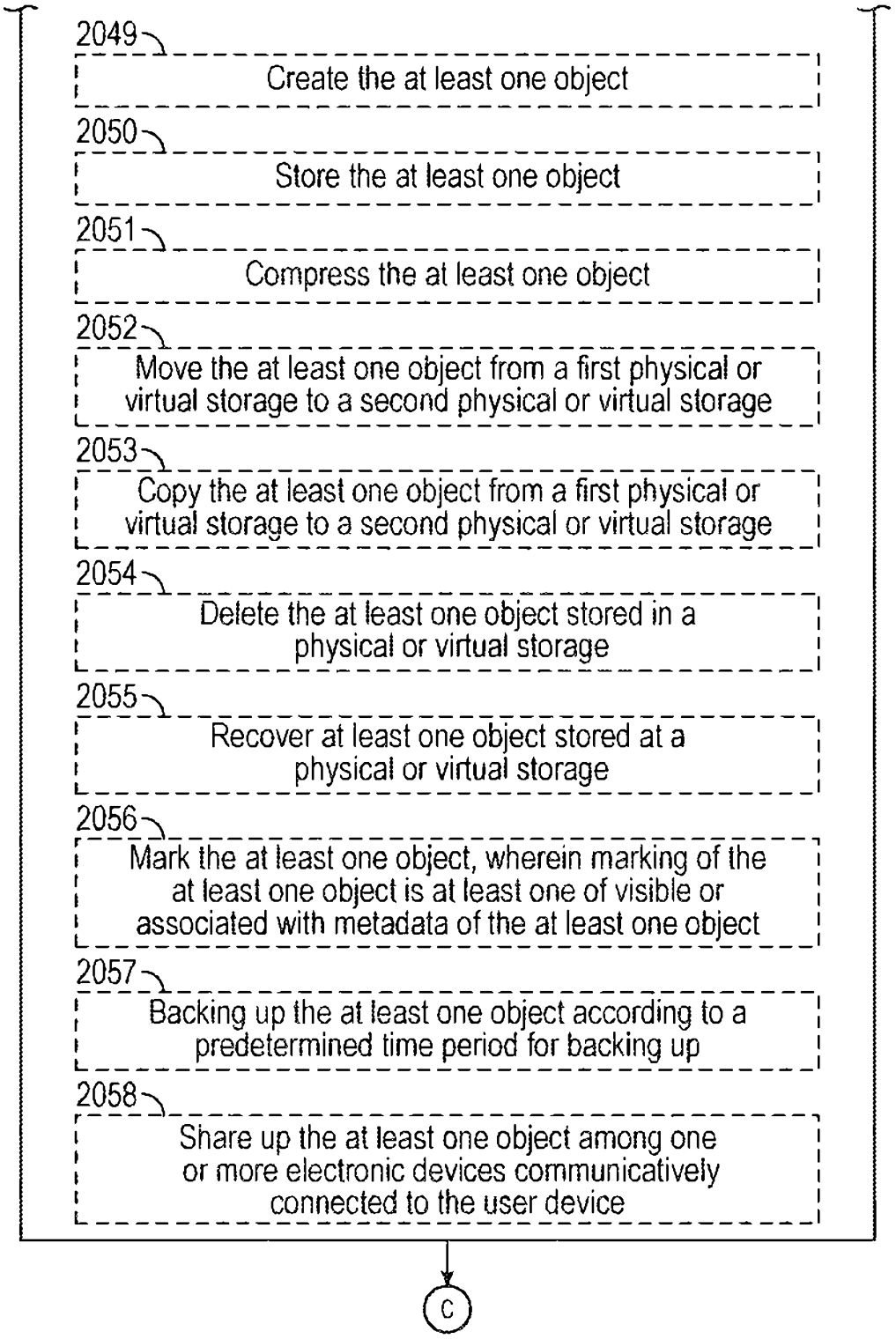

2049
Create the at least one object
2050
Store the at least one object
2051
Compress the at least one object
2052
Move the at least one object from a first physical or virtual storage to a second physical or virtual storage
2053
Copy the at least one object from a first physical or virtual storage to a second physical or virtual storage
2054
Delete the at least one object stored in a physical or virtual storage
2055
Recover at least one object stored at a physical or virtual storage
2056
Mark the at least one object, wherein marking of the at least one object is at least one of visible or associated with metadata of the at least one object
2057
Backing up the at least one object according to a predetermined time period for backing up
2058
Share up the at least one object among one or more electronic devices communicatively connected to the user device
C

FIG. 20E

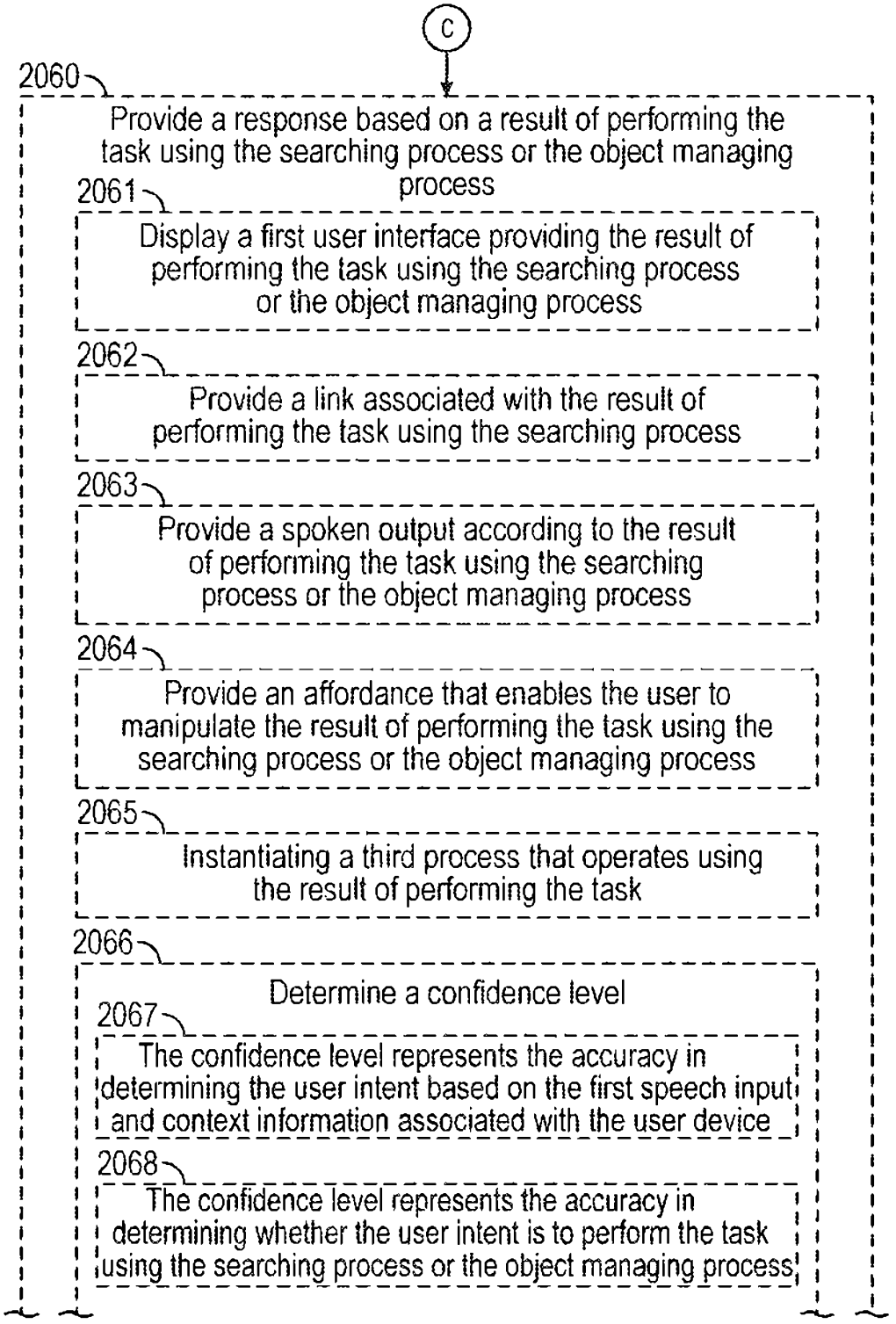

C
2060
Provide a response based on a result of performing the task using the searching process or the object managing process
2061
Display a first user interface providing the result of performing the task using the searching process or the object managing process
2062
Provide a link associated with the result of performing the task using the searching process
2063
Provide a spoken output according to the result of performing the task using the searching process or the object managing process
2064
Provide an affordance that enables the user to manipulate the result of performing the task using the searching process or the object managing process
2065
Instantiating a third process that operates using the result of performing the task
2066
Determine a confidence level
2067
The confidence level represents the accuracy in determining the user intent based on the first speech input and context information associated with the user device
2068
The confidence level represents the accuracy in determining whether the user intent is to perform the task using the searching process or the object managing process

FIG. 20F

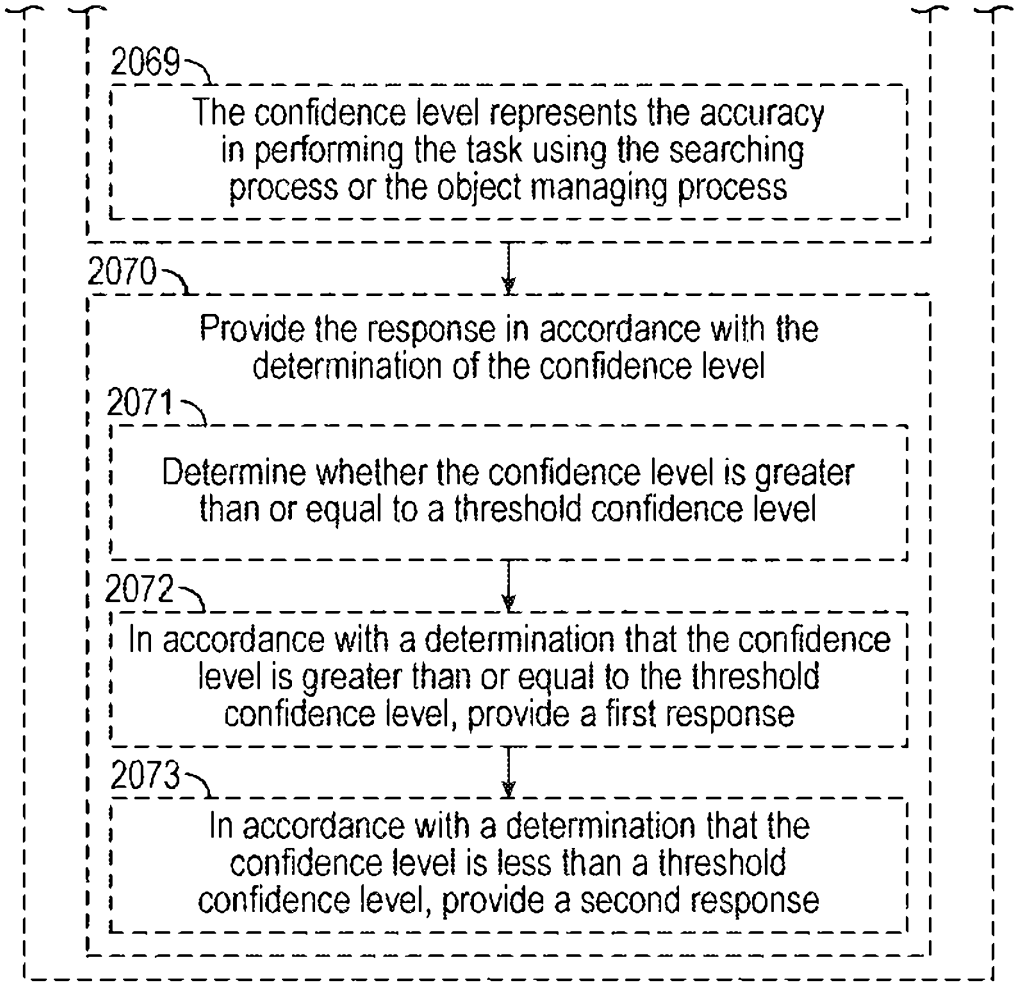

2069
The confidence level represents the accuracy in performing the task using the searching process or the object managing process
2070
Provide the response in accordance with the determination of the confidence level
2071
Determine whether the confidence level is greater than or equal to a threshold confidence level
2072
In accordance with a determination that the confidence level is greater than or equal to the threshold confidence level, provide a first response
2073
In accordance with a determination that the confidence level is less than a threshold confidence level, provide a second response

FIG. 20G

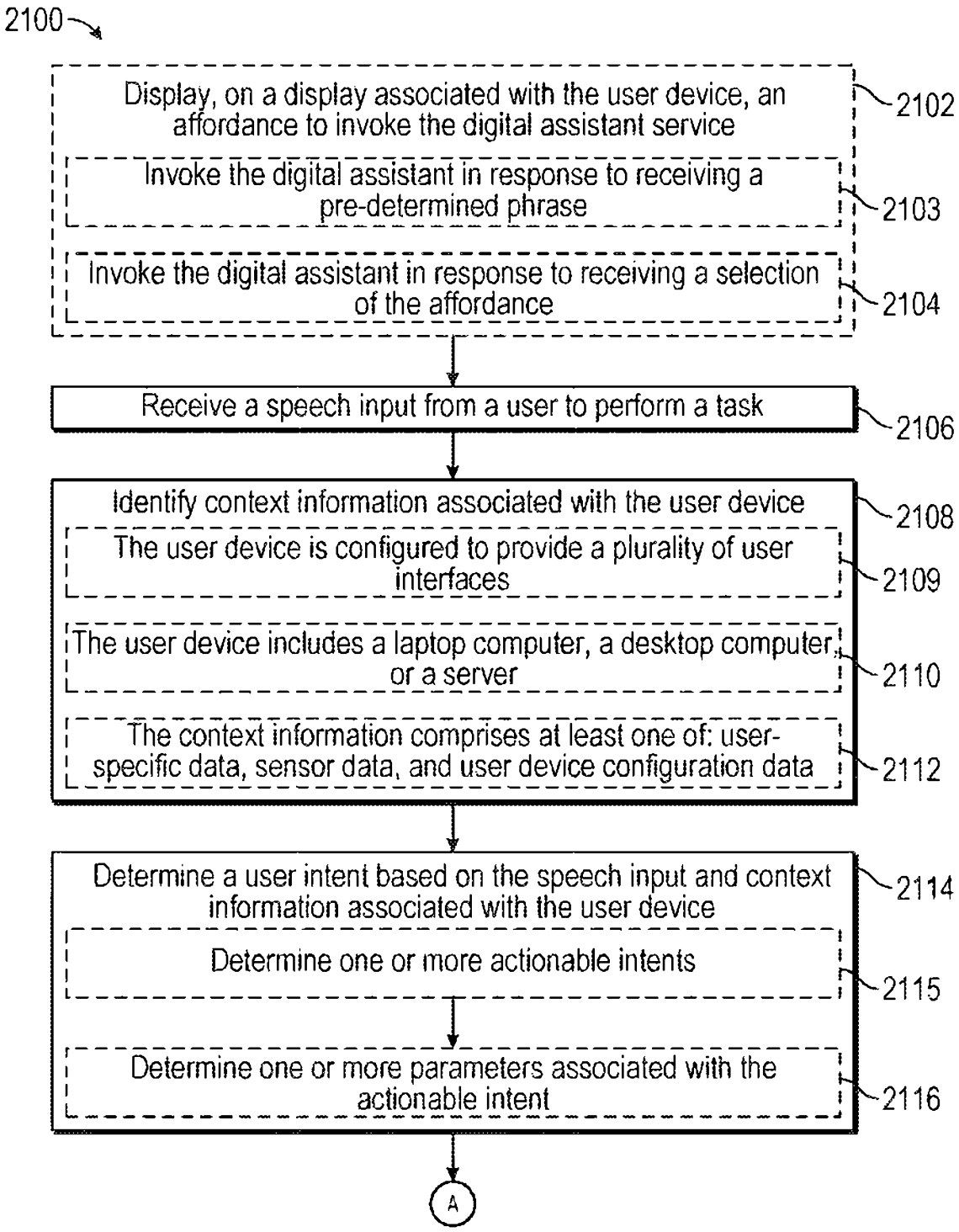

2100
Display, on a display associated with the user device, an affordance to invoke the digital assistant service
2102
Invoke the digital assistant in response to receiving a pre-determined phrase
2103
Invoke the digital assistant in response to receiving a selection of the affordance
2104
Receive a speech input from a user to perform a task
2106
Identify context information associated with the user device
2108
The user device is configured to provide a plurality of user interfaces
2109
The user device includes a laptop computer, a desktop computer, or a server
2110
The context information comprises at least one of: user-specific data, sensor data, and user device configuration data
2112
Determine a user intent based on the speech input and context information associated with the user device
2114
Determine one or more actionable intents
2115
Determine one or more parameters associated with the actionable intent
2116
A

FIG. 21A

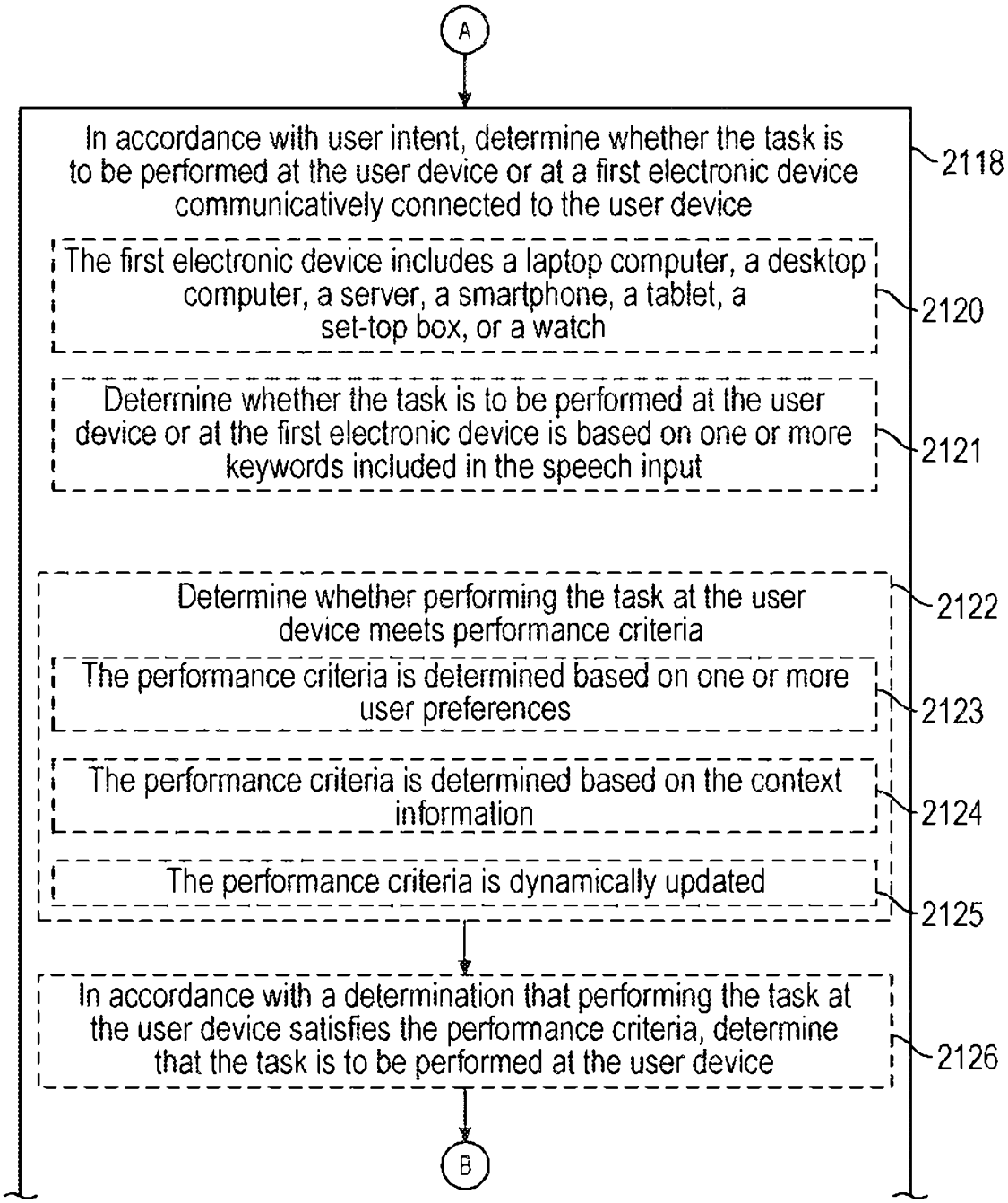

A
In accordance with user intent, determine whether the task is to be performed at the user device or at a first electronic device communicatively connected to the user device
2118
The first electronic device includes a laptop computer, a desktop computer, a server, a smartphone, a tablet, a set-top box, or a watch
2120
Determine whether the task is to be performed at the user device or at the first electronic device is based on one or more keywords included in the speech input
2121
Determine whether performing the task at the user device meets performance criteria
2122
The performance criteria is determined based on one or more user preferences
2123
The performance criteria is determined based on the context information
2124
The performance criteria is dynamically updated
2125
In accordance with a determination that performing the task at the user device satisfies the performance criteria, determine that the task is to be performed at the user device
2126
B

FIG. 21B

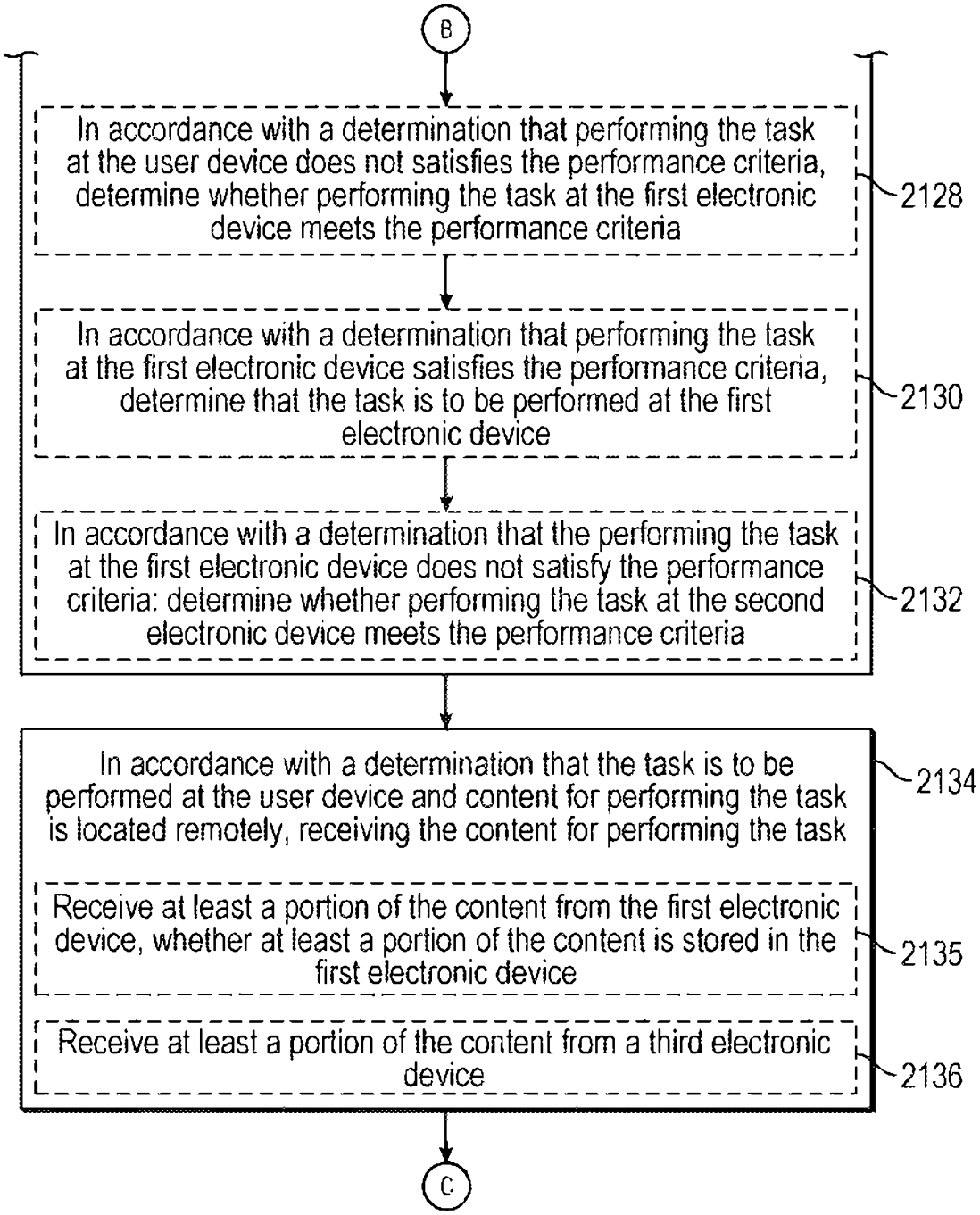

B
In accordance with a determination that performing the task at the user device does not satisfies the performance criteria, determine whether performing the task at the first electronic device meets the performance criteria
2128
In accordance with a determination that performing the task at the first electronic device satisfies the performance criteria, determine that the task is to be performed at the first electronic device
2130
In accordance with a determination that the performing the task at the first electronic device does not satisfy the performance criteria: determine whether performing the task at the second electronic device meets the performance criteria
2132
In accordance with a determination that the task is to be performed at the user device and content for performing the task is located remotely, receiving the content for performing the task
2134
Receive at least a portion of the content from the first electronic device, whether at least a portion of the content is stored in the first electronic device
2135
Receive at least a portion of the content from a third electronic device
2136
C

FIG. 21C

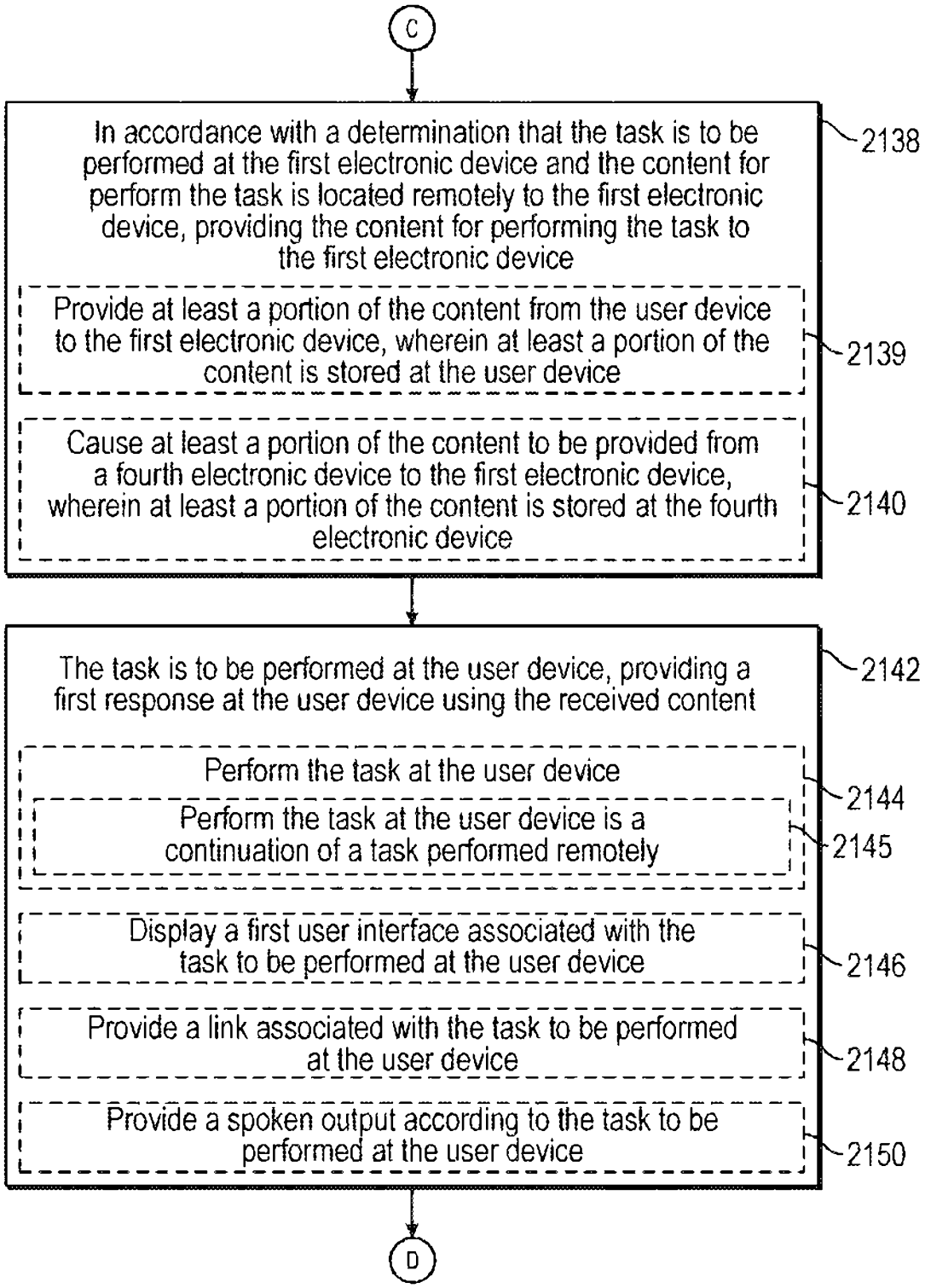

C
In accordance with a determination that the task is to be performed at the first electronic device and the content for perform the task is located remotely to the first electronic device, providing the content for performing the task to the first electronic device
2138
Provide at least a portion of the content from the user device to the first electronic device, wherein at least a portion of the content is stored at the user device
2139
Cause at least a portion of the content to be provided from a fourth electronic device to the first electronic device, wherein at least a portion of the content is stored at the fourth electronic device
2140
The task is to be performed at the user device, providing a first response at the user device using the received content
2142
Perform the task at the user device
2144
Perform the task at the user device is a continuation of a task performed remotely
2145
Display a first user interface associated with the task to be performed at the user device
2146
Provide a link associated with the task to be performed at the user device
2148
Provide a spoken output according to the task to be performed at the user device
2150
D

FIG. 21D

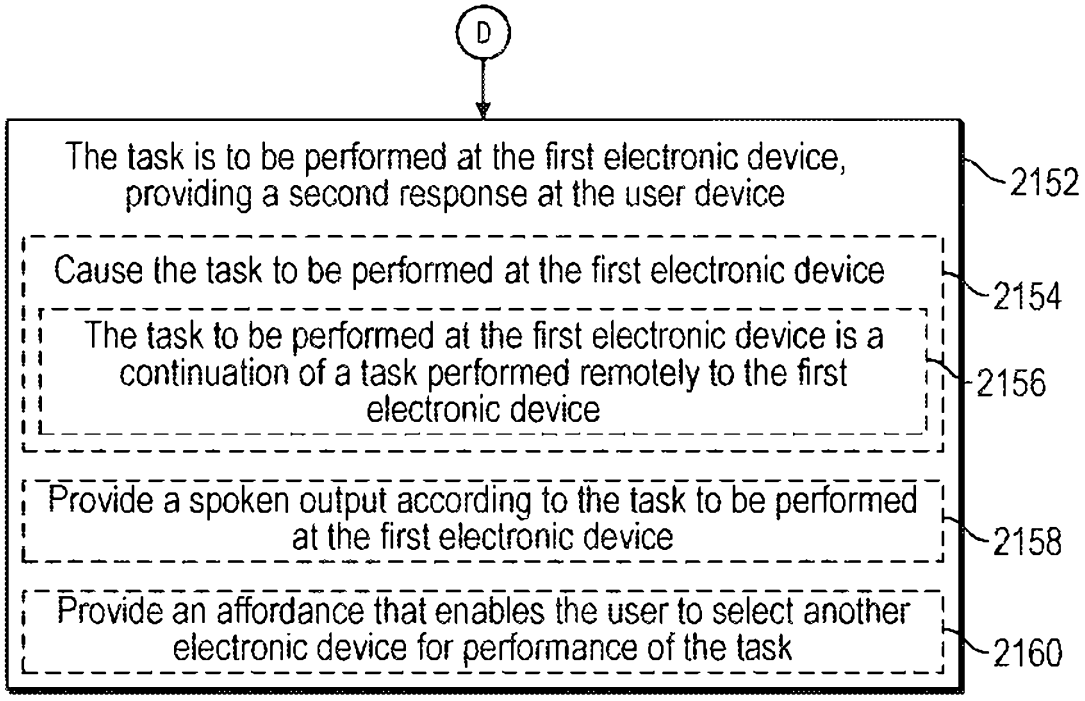

D
The task is to be performed at the first electronic device, providing a second response at the user device
2152
Cause the task to be performed at the first electronic device
2154
The task to be performed at the first electronic device is a continuation of a task performed remotely to the first electronic device
2156
Provide a spoken output according to the task to be performed at the first electronic device
2158
Provide an affordance that enables the user to select another electronic device for performance of the task
2160

FIG. 21E

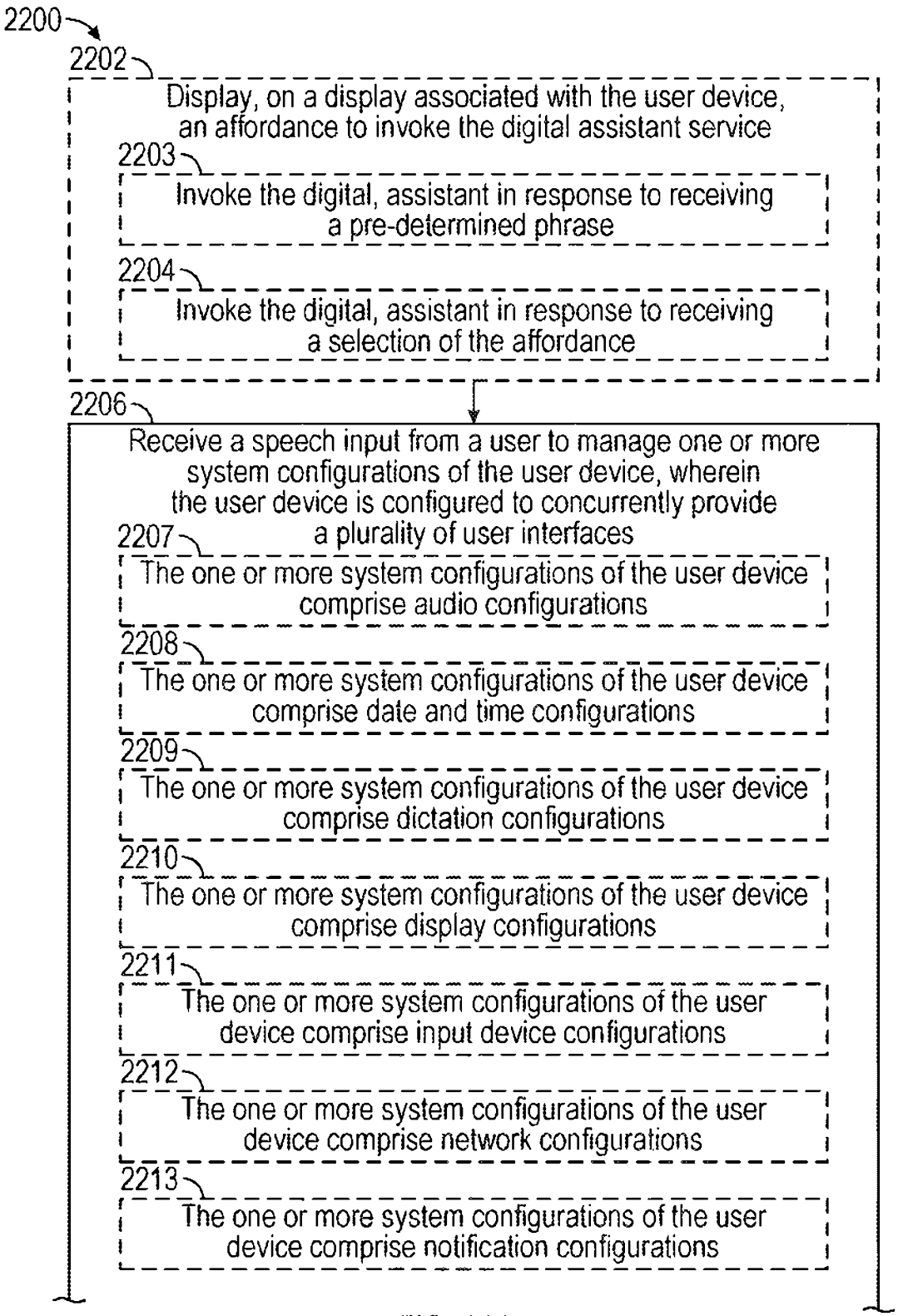

2200
2202
Display, on a display associated with the user device, an affordance to invoke the digital assistant service
2203
Invoke the digital, assistant in response to receiving a pre-determined phrase
2204
Invoke the digital, assistant in response to receiving a selection of the affordance
2206
Receive a speech input from a user to manage one or more system configurations of the user device, wherein the user device is configured to concurrently provide a plurality of user interfaces
2207
The one or more system configurations of the user device comprise audio configurations
2208
The one or more system configurations of the user device comprise date and time configurations
2209
The one or more system configurations of the user device comprise dictation configurations
2210
The one or more system configurations of the user device comprise display configurations
2211
The one or more system configurations of the user device comprise input device configurations
2212
The one or more system configurations of the user device comprise network configurations
2213
The one or more system configurations of the user device comprise notification configurations

FIG. 22A

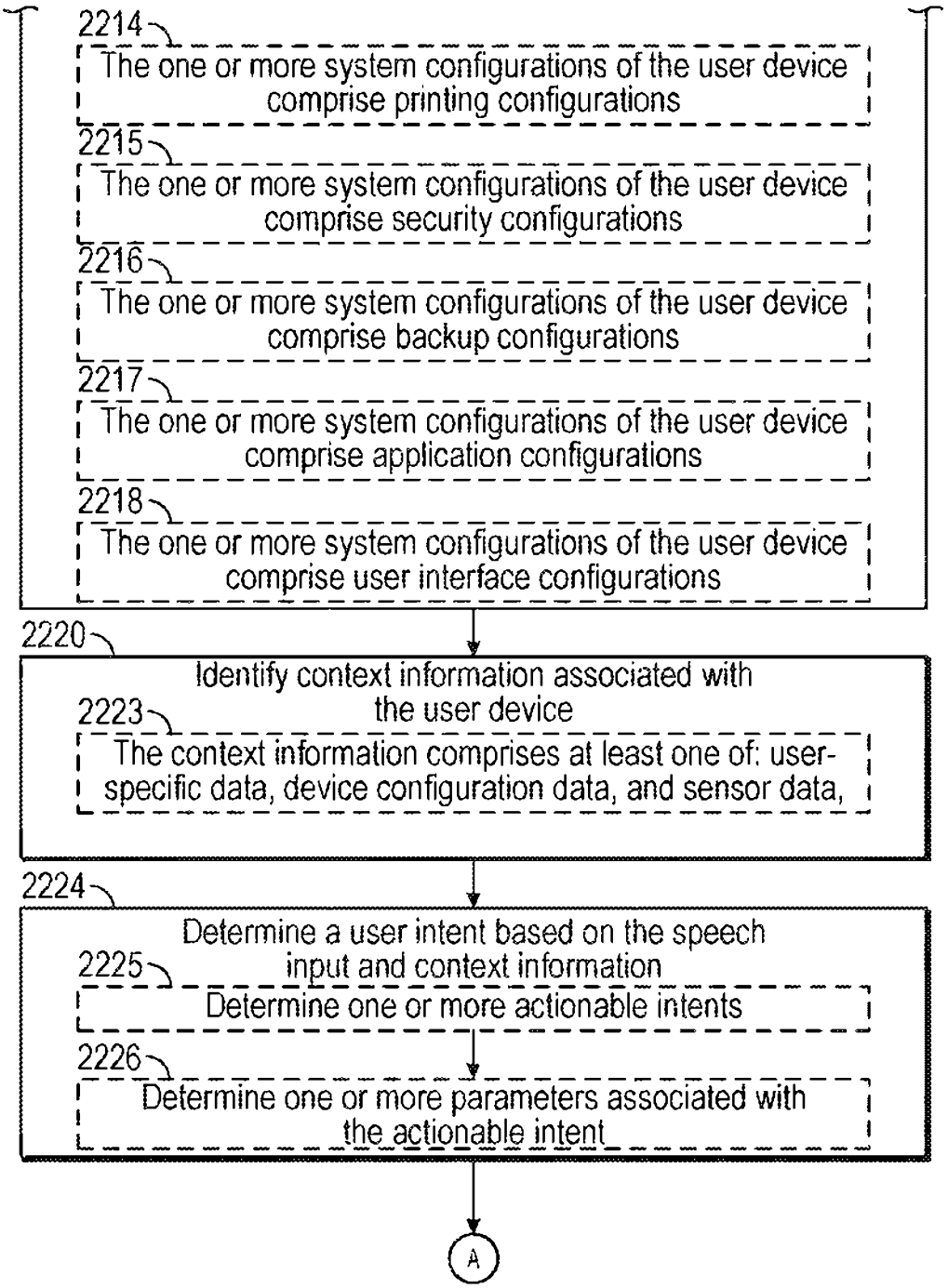

2214
The one or more system configurations of the user device comprise printing configurations
2215
The one or more system configurations of the user device comprise security configurations
2216
The one or more system configurations of the user device comprise backup configurations
2217
The one or more system configurations of the user device comprise application configurations
2218
The one or more system configurations of the user device comprise user interface configurations
2220
Identify context information associated with the user device
2223
The context information comprises at least one of: user-specific data, device configuration data, and sensor data,
2224
Determine a user intent based on the speech input and context information
2225
Determine one or more actionable intents
2226
Determine one or more parameters associated with the actionable intent
A

FIG. 22B

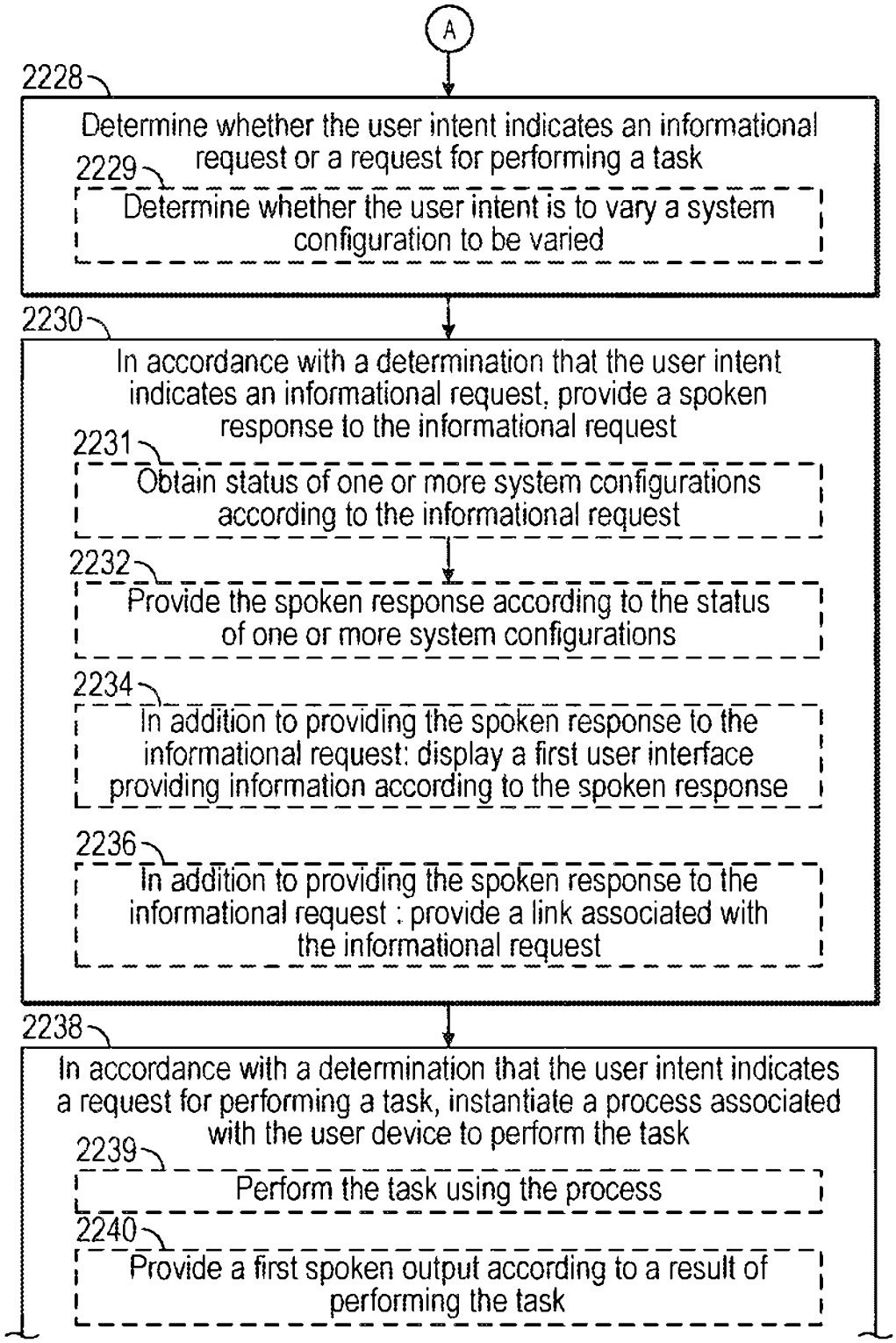

A
2228
Determine whether the user intent indicates an informational request or a request for performing a task
2229
Determine whether the user intent is to vary a system configuration to be varied
2230
In accordance with a determination that the user intent indicates an informational request, provide a spoken response to the informational request
2231
Obtain status of one or more system configurations according to the informational request
2232
Provide the spoken response according to the status of one or more system configurations
2234
In addition to providing the spoken response to the informational request: display a first user interface providing information according to the spoken response
2236
In addition to providing the spoken response to the informational request : provide a link associated with the informational request
2238
In accordance with a determination that the user intent indicates a request for performing a task, instantiate a process associated with the user device to perform the task
2239
Perform the task using the process
2240
Provide a first spoken output according to a result of performing the task

FIG. 22C

DK 2017 70434 A1

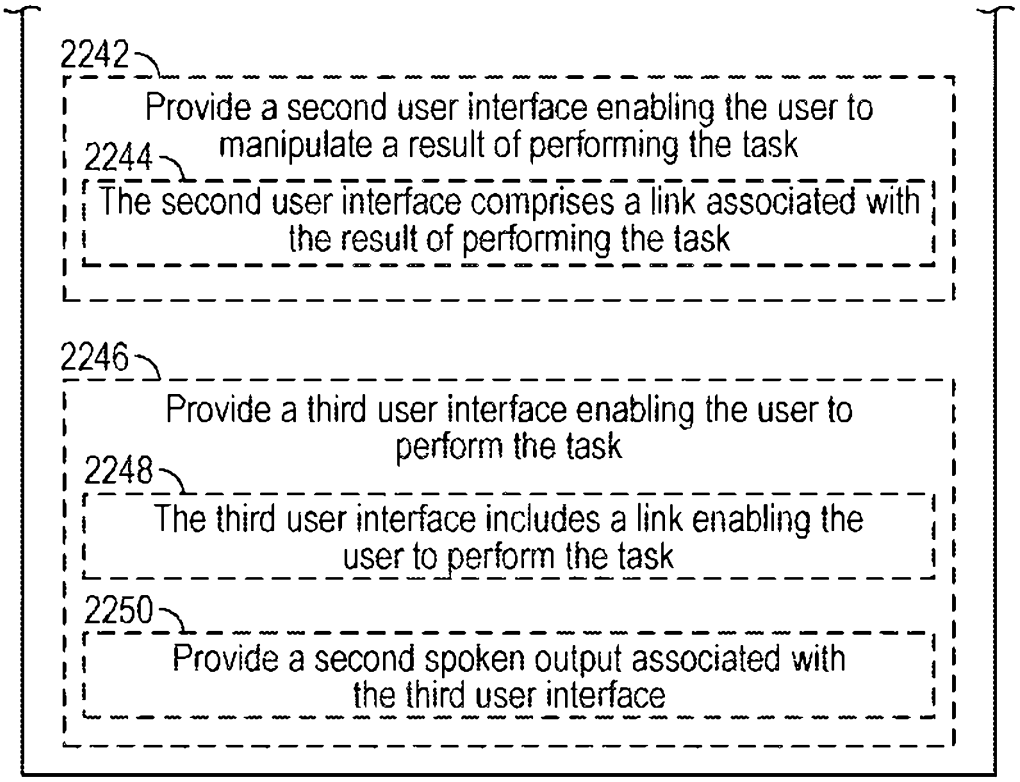


FIG. 22D

DK 2017 70434 A1

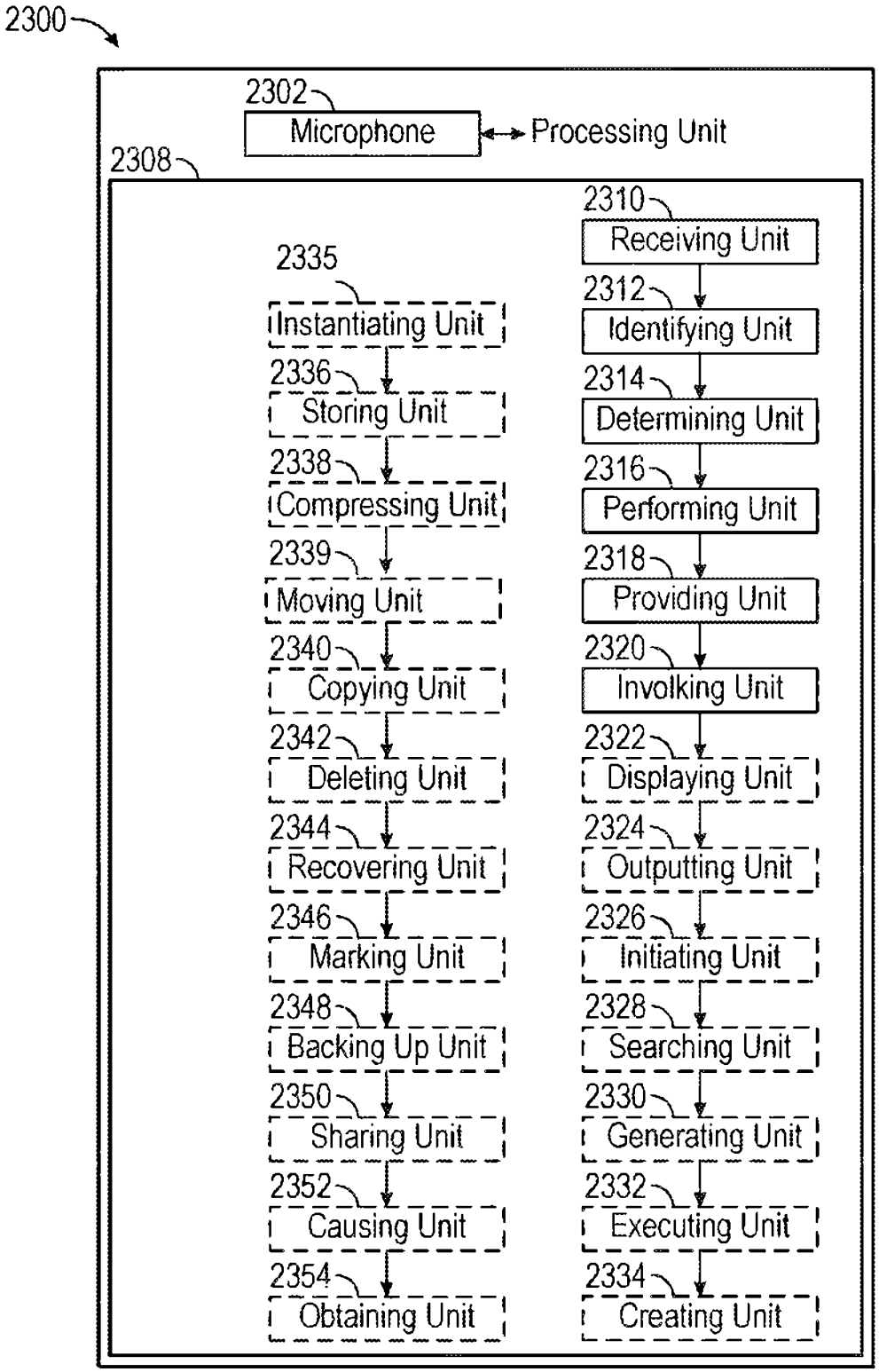


FIG. 23

PATENT APPLICATION IN THE U.S. PATENT AND TRADEMARK OFFICE

FOR

METHODS AND INTERFACES FOR HOME MEDIA CONTROL

FIELD

[0001] The present disclosure relates generally to computer user interfaces, and more specifically to techniques for managing media playback devices.

BACKGROUND

[0002] The number of electronic devices, and particularly smart devices, in users' homes continues to increase. These devices are increasingly being interconnected with each other, are increasingly more capable, and are performing more complex tasks. As such, these devices are increasingly expected to have thoughtfully-designed user interfaces.

BRIEF SUMMARY

[0003] Some techniques for managing media playback devices using electronic devices, however, are generally cumbersome and inefficient. For example, some existing techniques use a complex and time-consuming user interface, which may include multiple key presses or keystrokes. Existing techniques require more time than necessary, wasting user time and device energy. This latter consideration is particularly important in battery-operated devices.

[0004] Accordingly, the present technique provides electronic devices with faster, more efficient methods and interfaces for managing media playback devices. Such methods and interfaces optionally complement or replace other methods for managing media playback devices. Such methods and interfaces reduce the cognitive burden on a user and produce a more efficient human-machine interface. For battery-operated computing devices, such methods and interfaces conserve power and increase the time between battery charges. Such methods and interfaces can reduce the number of unnecessary, extraneous, or repetitive user inputs. Further, such methods and interfaces can reduce the battery usage by a display or speaker of an electronic device.

[0005] In accordance with some embodiments, a method is performed at a device with a display. The method comprises: receiving a user input; and in response to receiving the user input, displaying, on the display, a multi-device interface that includes: one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and a media playback status of the plurality of available playback devices.

[0006] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for: receiving a user input; and in response to receiving the user input, displaying, on the display, a multi-device interface that includes: one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and a media playback status of the plurality of available playback devices.

[0007] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: receiving a user input; and in response to receiving the user input, displaying, on the display, a multi-device interface that includes: one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and a media playback status of the plurality of available playback devices.

[0008] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; means for receiving a user input; and means for, responsive to receiving the user input, displaying, on the display, a multi-device interface that includes: one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and a media playback status of the plurality of available playback devices.

[0009] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for: receiving a user input; and in response to receiving the user input, displaying, on the display, a multi-device interface that includes: one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and a media playback status of the plurality of available playback devices.

[0010] In accordance with some embodiments, a method is performed at a device with a display. The method comprises: displaying, on the display, a first current media interface for the device, wherein the first current media interface includes: a representation of a first media item available for playback on the device, and a playback device selection affordance; receiving a first user input representing selection of the playback device selection affordance; in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device; receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and in response to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

[0011] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for: displaying, on the display, a first current media interface for the device, wherein the first current media interface includes: a representation of a first media item available for playback on the device, and a playback device selection affordance; receiving a first user input representing selection of the playback device selection affordance; in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device; receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and in response to receiving the

second user input, configuring the device to initiate playback on the plurality of available playback devices.

[0012] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: displaying, on the display, a first current media interface for the device, wherein the first current media interface includes: a representation of a first media item available for playback on the device, and a playback device selection affordance; receiving a first user input representing selection of the playback device selection affordance; in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device; receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and in response to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

[0013] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; means for displaying, on the display, a first current media interface for the device, wherein the first current media interface includes: a representation of a first media item available for playback on the device, and a playback device selection affordance; means for receiving a first user input representing selection of the playback device selection affordance; means for, responsive to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device; means for receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and means for, responsive to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

[0014] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for: displaying, on the display, a first

current media interface for the device, wherein the first current media interface includes: a representation of a first media item available for playback on the device, and a playback device selection affordance; receiving a first user input representing selection of the playback device selection affordance; in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device; receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and in response to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

[0015] In accordance with some embodiments, a method is performed at a first device. The method comprises: detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition, wherein the first device is connected to the second device via a communication link; in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link; subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received; in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

[0016] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for: detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition, wherein the first device is connected to the second device via a communication link; in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link; subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received; in accordance with a determination that the response audio tone signal from the second device has been

received, proceeding to a next step in a configuration process; and in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

[0017] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition, wherein the first device is connected to the second device via a communication link; in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link; subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received; in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

[0018] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition, wherein the first device is connected to the second device via a communication link; means for, in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link; means for, subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received; means for, in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and means for, in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

[0019] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for: detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition, wherein the first device is connected to the second device via a communication link; in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link; subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received; in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

[0020] In accordance with some embodiments, a method is performed at a first device with a display. The method comprises: detecting, while connected to a second device, an indication that a physical proximity between the first device and the second device satisfies a proximity condition; and in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

[0021] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, wherein the electronic device is a first device, the one or more programs including instructions for: detecting, while connected to a second device, an indication that a physical proximity between the first device and the second device satisfies a proximity condition; and in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

[0022] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs

including instructions for: detecting, while connected to a second device, an indication that a physical proximity between the electronic device and the second device satisfies a proximity condition; and in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

[0023] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; means for detecting, while connected to a second device, an indication that a physical proximity between the electronic device and the second device satisfies a proximity condition; and means for in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

[0024] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, wherein the electronic device is a first device, the one or more programs including instructions for: detecting, while connected to a second device, an indication that a physical proximity between the first device and the second device satisfies a proximity condition; and in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

[0025] In accordance with some embodiments, a method is performed at a device with a touch-sensitive surface. The method comprises: detecting a touch on the touch-sensitive surface; in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media; in accordance with a determination that the device is currently playing media: determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping; in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and in accordance with a determination that the device is not currently playing media, initiating playback of media.

[0026] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a touch-sensitive surface, the one or more programs including instructions for: detecting a touch on the touch-sensitive surface; in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media; in accordance with a determination that the device is currently playing media: determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping; in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and in accordance with a determination that the device is not currently playing media, initiating playback of media.

[0027] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a touch-sensitive surface; one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: detecting a touch on the touch-sensitive surface; in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media; in accordance with a determination that the device is currently playing media: determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping; in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and in accordance with a determination that the device is not currently playing media, initiating playback of media.

[0028] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a touch-sensitive surface; means for detecting a touch on the touch-sensitive surface; means for, responsive to detecting the touch on the touch-sensitive surface,

determining whether the device is currently playing media; means for, in accordance with a determination that the device is currently playing media: determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping; in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and means for, in accordance with a determination that the device is not currently playing media, initiating playback of media.

[0029] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a touch-sensitive surface, the one or more programs including instructions for: detecting a touch on the touch-sensitive surface; in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media; in accordance with a determination that the device is currently playing media: determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping; in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and in accordance with a determination that the device is not currently playing media, initiating playback of media.

[0030] In accordance with some embodiments, a method is performed at a first device. The method comprises: receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device; adding the first media item to a playback queue; receiving a second request to queue a second media item for playback on the set of devices; adding the second media item to the playback queue; accessing the first media item from a first media streaming service; initiating playback of the first media item on the set of devices; accessing the second media item from a second media streaming service different than

the first media streaming service; and initiating playback of the second media item on the set of devices.

[0031] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for: receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device; adding the first media item to a playback queue; receiving a second request to queue a second media item for playback on the set of devices; adding the second media item to the playback queue; accessing the first media item from a first media streaming service; initiating playback of the first media item on the set of devices; accessing the second media item from a second media streaming service different than the first media streaming service; and initiating playback of the second media item on the set of devices.

[0032] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device; adding the first media item to a playback queue; receiving a second request to queue a second media item for playback on the set of devices; adding the second media item to the playback queue; accessing the first media item from a first media streaming service; initiating playback of the first media item on the set of devices; accessing the second media item from a second media streaming service different than the first media streaming service; and initiating playback of the second media item on the set of devices.

[0033] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device; means for adding the first media item to a playback queue; means for receiving a second request to queue a second media item for playback on the set of devices; means for adding the second media item to

the playback queue; means for accessing the first media item from a first media streaming service; means for initiating playback of the first media item on the set of devices; means for accessing the second media item from a second media streaming service different than the first media streaming service; and means for initiating playback of the second media item on the set of devices.

[0034] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for: receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device; adding the first media item to a playback queue; receiving a second request to queue a second media item for playback on the set of devices; adding the second media item to the playback queue; accessing the first media item from a first media streaming service; initiating playback of the first media item on the set of devices; accessing the second media item from a second media streaming service different than the first media streaming service; and initiating playback of the second media item on the set of devices.

[0035] In accordance with some embodiments, a method is performed at a first device. The method comprises: receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device; adding the first media item to a playback queue; receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first; adding the second media item to the playback queue; initiating playback of the first media item on the set of devices; and initiating playback of the second media item on the set of devices.

[0036] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for: receiving, from a second device associated with a first user, a first request to queue a first media

item for playback on a set of devices, wherein the set of devices includes the first device; adding the first media item to a playback queue; receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first; adding the second media item to the playback queue; initiating playback of the first media item on the set of devices; and initiating playback of the second media item on the set of devices.

[0037] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device; adding the first media item to a playback queue; receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first; adding the second media item to the playback queue; initiating playback of the first media item on the set of devices; and initiating playback of the second media item on the set of devices.

[0038] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device; means for adding the first media item to a playback queue; means for receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first; means for adding the second media item to the playback queue; means for initiating playback of the first media item on the set of devices; and means for initiating playback of the second media item on the set of devices.

[0039] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for:

receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device; adding the first media item to a playback queue; receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first; adding the second media item to the playback queue; initiating playback of the first media item on the set of devices; and initiating playback of the second media item on the set of devices.

[0040] In accordance with some embodiments, a method is performed at a first device with a display. The method comprises: detecting an audio tone signal indicative of a pairing handshake with a second device, wherein the first device is connected to the second device via a communication link, and wherein the audio tone signal is out-of-band to the communication link; in accordance with detecting the audio tone signal, initiating, on the first device, a configuration process for configuring the second device; during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location; receiving user input selection of a location; and in response to receiving the user input, configuring the second device to be associated with the location.

[0041] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, wherein the electronic device is a first device, the one or more programs including instructions for: detecting an audio tone signal indicative of a pairing handshake with a second device, wherein the first device is connected to the second device via a communication link, and wherein the audio tone signal is out-of-band to the communication link; in accordance with detecting the audio tone signal, initiating, on the first device, a configuration process for configuring the second device; during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location; receiving user input selection of a location; and in response to receiving the user input, configuring the second device to be associated with the location.

[0042] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: detecting an audio tone signal indicative of a pairing handshake with a second device, wherein the electronic device is connected to the second device via a communication link, and wherein the audio tone signal is out-of-band to the communication link; in accordance with detecting the audio tone signal, initiating, on the electronic device, a configuration process for configuring the second device; during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location; receiving user input selection of a location; and in response to receiving the user input, configuring the second device to be associated with the location.

[0043] In accordance with some embodiments, an electronic device is described. The electronic device comprises: a display; means for detecting an audio tone signal indicative of a pairing handshake with a second device, wherein the electronic device is connected to the second device via a communication link, and wherein the audio tone signal is out-of-band to the communication link; means for, in accordance with detecting the audio tone signal, initiating, on the electronic device, a configuration process for configuring the second device; means for, during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location; means for receiving user input selection of a location; and means for, responsive to receiving the user input, configuring the second device to be associated with the location.

[0044] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device with a display, wherein the electronic device is a first device, the one or more programs including instructions for: detecting an audio tone signal indicative of a pairing handshake with a second device, wherein the first device is connected to the second device via a communication link, and wherein the audio tone signal is out-of-band to the communication link; in accordance with detecting the audio tone signal, initiating, on the first device, a configuration process for configuring the second device; during the configuration process, displaying, on the display, a

configuration user interface that includes a prompt to select a location; receiving user input selection of a location; and in response to receiving the user input, configuring the second device to be associated with the location.

[0045] In accordance with some embodiments, a method is performed at a device. The method comprises: while outputting an audible signal from the device, detecting a user input voice command from a voice command source; in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

[0046] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for: while outputting an audible signal from the device, detecting a user input voice command from a voice command source; in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

[0047] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: while outputting an audible signal from the device, detecting a user input voice command from a voice command source; in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

[0048] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for, while outputting an audible signal from the device, detecting a user input voice command from a voice command source; means for, responsive to detecting the user input voice command, determining a spatial position of the voice command

source relative to the device; and means for, while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

[0049] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for: while outputting an audible signal from the device, detecting a user input voice command from a voice command source; in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

[0050] In accordance with some embodiments, a method is performed at a first device associated with a first user account. The method comprises: detecting that a second device associated with a second user account is connected to the first device, wherein the first user account is associated with a first media preference profile; wherein the second user account is associated with a second media preference profile different than the first media preference profile; and initiating playback of media based on the first media preference profile and the second media preference profile.

[0051] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device associated with a first user account, wherein the first user account is associated with a first media preference profile, the one or more programs including instructions for: detecting that a second device associated with a second user account is connected to the first device, wherein the second user account is associated with a second media preference profile different than the first media preference profile; and initiating playback of media based on the first media preference profile and the second media preference profile.

[0052] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs

configured to be executed by the one or more processors, the one or more programs including instructions for: detecting that a second device associated with a second user account is connected to the electronic device, wherein the electronic device is associated with a first user account, wherein the first user account is associated with a first media preference profile: wherein the second user account is associated with a second media preference profile different than the first media preference profile; and initiating playback of media based on the first media preference profile and the second media preference profile.

[0053] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for detecting that a second device associated with a second user account is connected to the electronic device, wherein the electronic device is associated with a first user account, wherein the first user account is associated with a first media preference profile: wherein the second user account is associated with a second media preference profile different than the first media preference profile; and means for initiating playback of media based on the first media preference profile and the second media preference profile.

[0054] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device associated with a first user account, wherein the first user account is associated with a first media preference profile, the one or more programs including instructions for: detecting that a second device associated with a second user account is connected to the first device, wherein the second user account is associated with a second media preference profile different than the first media preference profile; and initiating playback of media based on the first media preference profile and the second media preference profile.

[0055] In accordance with some embodiments, a method is performed at a first device. The method comprises: receiving a request, from a third device, to initiate media playback on at least one device of a set of devices, wherein the set of devices comprises the first device and a second device, wherein the third device is not a member of the set of devices, and wherein the third device is connected to the first device; in response to receiving the request: in accordance with a determination that the third device has permission to initiate media playback on the at least one

device of set of devices, initiating media playback on the set of devices; and in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices: transmitting a request for permission to the second device; and in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

[0056] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for: receiving a request, from a third device, to initiate media playback on at least one device of a set of devices, wherein the set of devices comprises the first device and a second device, wherein the third device is not a member of the set of devices, and wherein the third device is connected to the first device; in response to receiving the request: in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices: transmitting a request for permission to the second device; and in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

[0057] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: receiving a request, from a third device, to initiate media playback on at least one device of a set of devices, wherein the set of devices comprises the electronic device and a second device, wherein the third device is not a member of the set of devices, and wherein the third device is connected to the electronic device; in response to receiving the request: in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices: transmitting a request for permission to the

second device; and in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

[0058] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for receiving a request, from a third device, to initiate media playback on at least one device of a set of devices, wherein the set of devices comprises the electronic device and a second device, wherein the third device is not a member of the set of devices, and wherein the third device is connected to the electronic device; means for, responsive to receiving the request: in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices: transmitting a request for permission to the second device; and in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

[0059] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for: receiving a request, from a third device, to initiate media playback on at least one device of a set of devices, wherein the set of devices comprises the first device and a second device, wherein the third device is not a member of the set of devices, and wherein the third device is connected to the first device; in response to receiving the request: in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices: transmitting a request for permission to the second device; and in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

[0060] In accordance with some embodiments, a method is performed at a first device associated with a user account. The method comprises: receiving user input representing a

request to access data from the user account; in response to receiving the user input, determining whether a physical positioning of the first device and a second device associated with the user account satisfies a permissive condition; in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the first device; and in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the first device.

[0061] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first electronic device associated with a user account, the one or more programs including instructions for: receiving user input representing a request to access data from the user account; in response to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition; in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

[0062] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: receiving user input representing a request to access data from a user account, wherein the electronic device is associated with the user account; in response to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition; in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

[0063] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for receiving user input representing a request to access data from a user account, wherein the electronic device is associated with the user account; means for, responsive to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition; means for, in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and means for, in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

[0064] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first electronic device associated with a user account, the one or more programs including instructions for: receiving user input representing a request to access data from the user account; in response to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition; in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

[0065] In accordance with some embodiments, a method is performed at a first device. The method comprises: receiving a command to initiate playback of media on a set of devices comprising the first device and a second device; in response to receiving the command, initiating synchronized playback of the media on the set of devices, wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and wherein the first device is configured to be a designated heir to the primary device.

[0066] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for: receiving a command to initiate playback of media on a set of devices comprising the first device and a second device; in response to receiving the command, initiating synchronized playback of the media on the set of devices, wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and wherein the first device is configured to be a designated heir to the primary device.

[0067] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for: receiving a command to initiate playback of media on a set of devices comprising the electronic device and a second device; in response to receiving the command, initiating synchronized playback of the media on the set of devices, wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and wherein the electronic device is configured to be a designated heir to the primary device.

[0068] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for receiving a command to initiate playback of media on a set of devices comprising the electronic device and a second device; means for, responsive to receiving the command, initiating synchronized playback of the media on the set of devices, wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and wherein the electronic device is configured to be a designated heir to the primary device.

[0069] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for:

receiving a command to initiate playback of media on a set of devices comprising the first device and a second device; in response to receiving the command, initiating synchronized playback of the media on the set of devices, wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and wherein the first device is configured to be a designated heir to the primary device.

[0070] In accordance with some embodiments, a method is performed at a first device of a set of devices associated with a first user account. The method comprises: receiving an audible output trigger event; determining that a first user, associated with the first user account, is located within an audible output range of a device of the set of devices; in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

[0071] In accordance with some embodiments, a non-transitory computer-readable storage medium is described. The non-transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device of a set of devices associated with a first user account, the one or more programs including instructions for: receiving an audible output trigger event; determining that a first user, associated with the first user account, is located within an audible output range of a device of the set of devices; in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

[0072] In accordance with some embodiments, an electronic device is described. The electronic device comprises: one or more processors; and memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including

instructions for: receiving an audible output trigger event; determining that a first user, associated with a first user account, is located within an audible output range of a device of a set of devices associated with the first user account, wherein the electronic device is a first device of the set of devices; in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

[0073] In accordance with some embodiments, an electronic device is described. The electronic device comprises: means for receiving an audible output trigger event; means for determining that a first user, associated with a first user account, is located within an audible output range of a device of a set of devices associated with the first user account, wherein the electronic device is a first device of the set of devices; means for, in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and means for, in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

[0074] In accordance with some embodiments, a transitory computer-readable storage medium is described. The transitory computer-readable storage medium stores one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device of a set of devices associated with a first user account, the one or more programs including instructions for: receiving an audible output trigger event; determining that a first user, associated with the first user account, is located within an audible output range of a device of the set of devices; in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

[0075] Executable instructions for performing these functions are, optionally, included in a non-transitory computer-readable storage medium or other computer program product configured for execution by one or more processors. Executable instructions for performing these functions are, optionally, included in a transitory computer-readable storage medium or other computer program product configured for execution by one or more processors.

[0076] Thus, devices are provided with faster, more efficient methods and interfaces for managing media playback devices, thereby increasing the effectiveness, efficiency, and user satisfaction with such devices. Such methods and interfaces may complement or replace other methods for managing media playback devices.

DESCRIPTION OF THE FIGURES

[0077] 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.

[0078] FIG. 1A is a block diagram illustrating a portable multifunction device with a touch-sensitive display in accordance with some embodiments.

[0079] FIG. 1B is a block diagram illustrating exemplary components for event handling in accordance with some embodiments.

[0080] FIG. 2 illustrates a portable multifunction device having a touch screen in accordance with some embodiments.

[0081] FIG. 3 is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface in accordance with some embodiments.

[0082] FIG. 4A illustrates an exemplary user interface for a menu of applications on a portable multifunction device in accordance with some embodiments.

[0083] FIG. 4B 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.

[0084] FIG. 5A illustrates a personal electronic device in accordance with some embodiments.

[0085] FIG. 5B is a block diagram illustrating a personal electronic device in accordance with some embodiments.

[0086] FIGS. 5C-5D illustrate exemplary components of a personal electronic device having a touch-sensitive display and intensity sensors in accordance with some embodiments.

[0087] FIGS. 5E-5H illustrate exemplary components and user interfaces of a personal electronic device in accordance with some embodiments.

[0088] FIG. 5I illustrates an electronic device in accordance with some embodiments.

[0089] FIG. 5J is a block diagram illustrating an electronic device in accordance with some embodiments.

[0090] FIGS. 6A-6AR illustrate exemplary techniques and user interfaces for playing back media in accordance with some embodiments.

[0091] FIG. 7 is a flow diagram illustrating a process for playing back media in accordance with some embodiments.

[0092] FIGS. 8A-8P illustrate exemplary techniques and user interfaces for playing back media in accordance with some embodiments.

[0093] FIG. 9 is a flow diagram illustrating a process for playing back media in accordance with some embodiments.

[0094] FIGS. 10A-10O illustrate exemplary techniques and user interfaces for in accordance with some embodiments.

[0095] FIG. 11 is a flow diagram illustrating a process for displaying content in accordance with some embodiments.

[0096] FIGS. 12A-12AM illustrate exemplary techniques and user interfaces for viewing media information in accordance with some embodiments.

[0097] FIG. 13 is a flow diagram illustrating a process for viewing media information in accordance with some embodiments.

[0098] FIGS. 14A-14T illustrate exemplary techniques and user interfaces for controlling an electronic device in accordance with some embodiments.

[0099] FIG. 15 is a flow diagram illustrating a process for controlling an electronic device in accordance with some embodiments.

[0100] FIGS. 16A-16P illustrate exemplary techniques and user interfaces for managing a playback queue in accordance with some embodiments.

[0101] FIG. 17 is a flow diagram illustrating a process for managing a playback queue in accordance with some embodiments.

[0102] FIG. 18 is a flow diagram illustrating a process for managing a playback queue in accordance with some embodiments.

[0103] FIGS. 19A-19P illustrate exemplary techniques and user interfaces for configuring an electronic device in accordance with some embodiments.

[0104] FIG.20 is a flow diagram illustrating a process for configuring an electronic device in accordance with some embodiments.

[0105] FIGS. 21A-21I illustrate exemplary techniques and user interfaces for adjusting an audible output of an electronic device in accordance with some embodiments.

[0106] FIG. 22 is a flow diagram illustrating a process for adjusting an audible output of an electronic device in accordance with some embodiments.

[0107] FIGS. 23A-23K illustrate exemplary techniques and user interfaces for managing a playback queue in accordance with some embodiments.

[0108] FIG. 24 is a flow diagram illustrating a process for managing a playback queue in accordance with some embodiments.

[0109] FIGS. 25A-25I illustrate exemplary techniques and user interfaces for playing back media in accordance with some embodiments.

[0110] FIG. 26 is a flow diagram illustrating a process for playing back media in accordance with some embodiments.

[0111] FIGS. 27A-27E illustrate exemplary techniques and user interfaces for providing access to an electronic device in accordance with some embodiments.

[0112] FIG. 28 is a flow diagram illustrating a process for providing access to an electronic device in accordance with some embodiments.

[0113] FIGS. 29A-29I illustrate exemplary techniques and user interfaces for managing media playback in accordance with some embodiments.

[0114] FIG. 30 is a flow diagram illustrating a process for managing media playback in accordance with some embodiments.

[0115] FIGS. 31A-31N illustrate exemplary techniques and user interfaces for providing output in accordance with some embodiments.

[0116] FIG. 32 is a flow diagram illustrating a process for providing output in accordance with some embodiments.

DESCRIPTION OF EMBODIMENTS

[0117] The following description sets forth exemplary methods, parameters, and the like. It should be recognized, however, that such description is not intended as a limitation on the scope of the present disclosure but is instead provided as a description of exemplary embodiments.

[0118] There is a need for electronic devices that provide efficient methods and interfaces for managing media playback devices. For example, methods and techniques for transferring playback between devices are described below. Such techniques can reduce the cognitive burden

on a user who accesses event notifications, thereby enhancing productivity. Further, such techniques can reduce processor and battery power otherwise wasted on redundant user inputs.

[0119] Below, FIGS. 1A-1B, 2, 3, 4A-4B, and 5A-5H provide a description of exemplary devices for performing the techniques for managing event notifications. FIGS. 6A-6AR illustrate exemplary techniques and user interfaces for playing back media in accordance with some embodiments. FIG. 7 is a flow diagram illustrating a process for playing back media in accordance with some embodiments. The user interfaces in FIGS. 6A-6AR are used to illustrate the processes described below, including the processes in FIGS. 7. FIGS. 8A-8P illustrate exemplary techniques and user interfaces for playing back media in accordance with some embodiments. FIG. 9 is a flow diagram illustrating a process for playing back media in accordance with some embodiments. The user interfaces in FIGS. 8A-8P are used to illustrate the processes described below, including the processes in FIGS. 9. FIGS. 10A-10O illustrate exemplary techniques and user interfaces for in accordance with some embodiments. FIG. 11 is a flow diagram illustrating a process for displaying content in accordance with some embodiments. The user interfaces in FIGS. 10A-10O are used to illustrate the processes described below, including the processes in FIGS. 11. FIGS. 12A-12AM illustrate exemplary techniques and user interfaces for viewing media information in accordance with some embodiments. FIG. 13 is a flow diagram illustrating a process for viewing media information in accordance with some embodiments. The user interfaces in FIGS. 12A-12AM are used to illustrate the processes described below, including the processes in FIGS. 13. FIGS. 14A-14T illustrate exemplary techniques and user interfaces for controlling an electronic device in accordance with some embodiments. FIG. 15 is a flow diagram illustrating a process for controlling an electronic device in accordance with some embodiments. The user interfaces in FIGS. 14A-14T are used to illustrate the processes described below, including the processes in FIGS. 15. FIGS. 16A-16P illustrate exemplary techniques and user interfaces for managing a playback queue in accordance with some embodiments. FIG. 17 is a flow diagram illustrating a process for managing a playback queue in accordance with some embodiments. FIG. 18 is a flow diagram illustrating a process for managing a playback queue in accordance with some embodiments. The user interfaces in FIGS. 16A-16P are used to illustrate the processes described below, including the processes in FIGS. 17 and 18. FIGS. 19A-19P illustrate exemplary techniques and user interfaces for configuring an electronic device in accordance with some embodiments. FIG. 20 is

a flow diagram illustrating a process for configuring an electronic device in accordance with some embodiments. The user interfaces in FIGS. 19A-19P are used to illustrate the processes described below, including the processes in FIGS. 20. FIGS. 21A-21I illustrate exemplary techniques and user interfaces for adjusting an audible output of an electronic device in accordance with some embodiments. FIG. 22 is a flow diagram illustrating a process for adjusting an audible output of an electronic device in accordance with some embodiments. The user interfaces in FIGS. 21A-21I are used to illustrate the processes described below, including the processes in FIGS. 22. FIGS. 23A-23K illustrate exemplary techniques and user interfaces for managing a playback queue in accordance with some embodiments. FIG. 24 is a flow diagram illustrating a process for managing a playback queue in accordance with some embodiments. The user interfaces in FIGS. 23A-23K are used to illustrate the processes described below, including the processes in FIGS. 24. FIGS. 25A-25I illustrate exemplary techniques and user interfaces for playing back media in accordance with some embodiments. FIG. 26 is a flow diagram illustrating a process for playing back media in accordance with some embodiments. The user interfaces in FIGS. 25A-25I are used to illustrate the processes described below, including the processes in FIGS. 26. FIGS. 27A-27E illustrate exemplary techniques and user interfaces for providing access to an electronic device in accordance with some embodiments. FIG. 28 is a flow diagram illustrating a process for providing access to an electronic device in accordance with some embodiments. The user interfaces in FIGS. 27A-27E are used to illustrate the processes described below, including the processes in FIGS. 28. FIGS. 29A-29I illustrate exemplary techniques and user interfaces for managing media playback in accordance with some embodiments. FIG. 30 is a flow diagram illustrating a process for managing media playback in accordance with some embodiments. The user interfaces in FIGS. 29A-29I are used to illustrate the processes described below, including the processes in FIGS. 30. FIGS. 31A-31N illustrate exemplary techniques and user interfaces for providing output in accordance with some embodiments. FIG. 32 is a flow diagram illustrating a process for providing output in accordance with some embodiments. The user interfaces in FIGS. 31A-31N are used to illustrate the processes described below, including the processes in FIGS. 32.

[0120] 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 touch could be termed a second

touch, and, similarly, a second touch could be termed a first touch, without departing from the scope of the various described embodiments. The first touch and the second touch are both touches, but they are not the same touch.

[0121] The terminology used in the description of the various described embodiments herein is for the purpose of describing particular embodiments only and is not intended to be limiting. As used in the description of the various described embodiments 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.

[0122] The term "if" is, optionally, 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" is, optionally, 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.

[0123] Embodiments of electronic devices, user interfaces for such devices, and associated processes for using such devices are described. In some embodiments, the device is a portable communications device, such as a mobile telephone, that also contains other functions, such as PDA and/or music player functions. Exemplary embodiments of portable multifunction devices include, without limitation, the iPhone®, iPod Touch®, and iPad® devices from Apple Inc. of Cupertino, California. Other portable electronic devices, such as laptops or tablet computers with touch-sensitive surfaces (e.g., touch screen displays and/or touchpads), are, optionally, used. It should also be understood that, in some embodiments, the device is not a portable communications device, but is a desktop computer with a touch-sensitive surface (e.g., a touch screen display and/or a touchpad).

[0124] In the discussion that follows, an electronic device that includes a display and a touch-sensitive surface is described. It should be understood, however, that the electronic device optionally includes one or more other physical user-interface devices, such as a physical keyboard, a mouse, and/or a joystick.

[0125] The device typically supports a variety of applications, such as one or more of the following: a drawing application, a presentation application, a word processing application, a website creation application, a disk authoring application, a spreadsheet application, a gaming application, a telephone application, a video conferencing application, an e-mail application, an instant messaging application, a workout support application, a photo management application, a digital camera application, a digital video camera application, a web browsing application, a digital music player application, and/or a digital video player application.

[0126] The various applications that are executed on the device optionally use at least one common physical user-interface device, such as the touch-sensitive surface. One or more functions of the touch-sensitive surface as well as corresponding information displayed on the device are, optionally, adjusted and/or varied from one application to the next and/or within a respective application. In this way, a common physical architecture (such as the touch-sensitive surface) of the device optionally supports the variety of applications with user interfaces that are intuitive and transparent to the user.

[0127] Attention is now directed toward embodiments of portable devices with touch-sensitive displays. FIG. 1A is a block diagram illustrating portable multifunction device 100 with touch-sensitive display system 112 in accordance with some embodiments. Touch-sensitive display 112 is sometimes called a "touch screen" for convenience and is sometimes known as or called a "touch-sensitive display system." Device 100 includes memory 102 (which optionally includes one or more computer-readable storage mediums), memory controller 122, one or more processing units (CPUs) 120, peripherals interface 118, RF circuitry 108, audio circuitry 110, speaker 111, microphone 113, input/output (I/O) subsystem 106, other input control devices 116, and external port 124. Device 100 optionally includes one or more optical sensors 164. Device 100 optionally includes one or more contact intensity sensors 165 for detecting intensity of contacts on device 100 (e.g., a touch-sensitive surface such as touch-sensitive display system 112

of device 100). Device 100 optionally includes one or more tactile output generators 167 for generating tactile outputs on device 100 (e.g., generating tactile outputs on a touch-sensitive surface such as touch-sensitive display system 112 of device 100 or touchpad 355 of device 300). These components optionally communicate over one or more communication buses or signal lines 103.

[0128] 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).

[0129] 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.

[0130] It should be appreciated that device 100 is only one example of a portable multifunction device, and that device 100 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. 1A 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.

[0131] Memory 102 optionally includes high-speed random access memory and optionally also includes 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 122 optionally controls access to memory 102 by other components of device 100.

[0132] Peripherals interface 118 can be used to couple input and output peripherals of the device to CPU 120 and memory 102. The one or more processors 120 run or execute various software programs and/or sets of instructions stored in memory 102 to perform various functions for device 100 and to process data. In some embodiments, peripherals interface 118, CPU 120, and memory controller 122 are, optionally, implemented on a single chip, such as chip 104. In some other embodiments, they are, optionally, implemented on separate chips.

[0133] RF (radio frequency) circuitry 108 receives and sends RF signals, also called electromagnetic signals. RF circuitry 108 converts electrical signals to/from electromagnetic signals and communicates with communications networks and other communications devices via the electromagnetic signals. RF circuitry 108 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 108 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 108 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.

[0134] Audio circuitry 110, speaker 111, and microphone 113 provide an audio interface between a user and device 100. Audio circuitry 110 receives audio data from peripherals interface 118, converts the audio data to an electrical signal, and transmits the electrical signal to speaker 111. Speaker 111 converts the electrical signal to human-audible sound waves. Audio circuitry 110 also receives electrical signals converted by microphone 113 from sound waves. Audio circuitry 110 converts the electrical signal to audio data and transmits the audio data to peripherals interface 118 for processing. Audio data is, optionally, retrieved from and/or transmitted to memory 102 and/or RF circuitry 108 by peripherals interface 118. In some embodiments, audio circuitry 110 also includes a headset jack (e.g., 212, FIG. 2). The headset jack provides an interface between audio circuitry 110 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).

[0135] I/O subsystem 106 couples input/output peripherals on device 100, such as touch screen 112 and other input control devices 116, to peripherals interface 118. I/O subsystem 106 optionally includes display controller 156, optical sensor controller 158, intensity sensor controller 159, haptic feedback controller 161, and one or more input controllers 160 for other input or control devices. The one or more input controllers 160 receive/send electrical signals from/to other input control devices 116. The other input control devices 116 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) 160 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., 208, FIG. 2) optionally include an up/down button for volume control of speaker 111 and/or microphone 113. The one or more buttons optionally include a push button (e.g., 206, FIG. 2).

[0136] A quick press of the push button optionally disengages a lock of touch screen 112 or optionally begins 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., 206) optionally turns power to device 100 on or off. The functionality of one or more of the buttons are, optionally, user-customizable. Touch screen 112 is used to implement virtual or soft buttons and one or more soft keyboards.

[0137] Touch-sensitive display 112 provides an input interface and an output interface between the device and a user. Display controller 156 receives and/or sends electrical signals from/to touch screen 112. Touch screen 112 displays visual output to the user. The visual output optionally includes graphics, text, icons, video, and any combination thereof (collectively termed "graphics"). In some embodiments, some or all of the visual output optionally corresponds to user-interface objects.

[0138] Touch screen 112 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 112 and display controller 156 (along with any associated modules and/or sets of instructions in memory 102) detect contact (and any movement or breaking of the contact) on touch screen 112 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 112. In an exemplary embodiment, a point of contact between touch screen 112 and the user corresponds to a finger of the user.

[0139] Touch screen 112 optionally uses LCD (liquid crystal display) technology, LPD (light emitting polymer display) technology, or LED (light emitting diode) technology, although other display technologies are used in other embodiments. Touch screen 112 and display controller

156 optionally 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 112. 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.

[0140] A touch-sensitive display in some embodiments of touch screen 112 is, optionally, 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 112 displays visual output from device 100, whereas touch-sensitive touchpads do not provide visual output.

[0141] A touch-sensitive display in some embodiments of touch screen 112 is 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.

[0142] Touch screen 112 optionally has a video resolution in excess of 100 dpi. In some embodiments, the touch screen has a video resolution of approximately 160 dpi. The user optionally makes contact with touch screen 112 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.

[0143] In some embodiments, in addition to the touch screen, device 100 optionally includes 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 is, optionally, a touch-sensitive surface that is separate from touch screen 112 or an extension of the touch-sensitive surface formed by the touch screen.

[0144] Device 100 also includes power system 162 for powering the various components. Power system 162 optionally includes 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.

[0145] Device 100 optionally also includes one or more optical sensors 164. FIG. 1A shows an optical sensor coupled to optical sensor controller 158 in I/O subsystem 106. Optical sensor 164 optionally includes charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) phototransistors. Optical sensor 164 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 143 (also called a camera module), optical sensor 164 optionally captures still images or video. In some embodiments, an optical sensor is located on the back of device 100, opposite touch screen display 112 on the front of the device so that the touch screen display is enabled for use 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 is, optionally, 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 164 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 164 is used along with the touch screen display for both video conferencing and still and/or video image acquisition.

[0146] Device 100 optionally also includes one or more contact intensity sensors 165. FIG. 1A shows a contact intensity sensor coupled to intensity sensor controller 159 in I/O subsystem 106. Contact intensity sensor 165 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 165 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 112). In some embodiments, at least one contact intensity sensor is located on the back of device 100, opposite touch screen display 112, which is located on the front of device 100.

[0147] Device 100 optionally also includes one or more proximity sensors 166. FIG. 1A shows proximity sensor 166 coupled to peripherals interface 118. Alternately, proximity sensor 166 is, optionally, coupled to input controller 160 in I/O subsystem 106. Proximity sensor 166 optionally performs 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 112 when the multifunction device is placed near the user's ear (e.g., when the user is making a phone call).

[0148] Device 100 optionally also includes one or more tactile output generators 167. FIG. 1A shows a tactile output generator coupled to haptic feedback controller 161 in I/O subsystem 106. Tactile output generator 167 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 165 receives tactile feedback generation instructions from haptic feedback module 133 and generates tactile outputs on device 100 that are capable of being sensed by a user of device 100. 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 112) and, optionally, generates a tactile output by moving the touch-sensitive surface vertically (e.g., in/out of a surface of device 100) or laterally (e.g., back and forth in the same plane as a surface of device 100). In some embodiments, at least one tactile output generator sensor is located on the back of device 100, opposite touch screen display 112, which is located on the front of device 100.

[0149] Device 100 optionally also includes one or more accelerometers 168. FIG. 1A shows accelerometer 168 coupled to peripherals interface 118. Alternately, accelerometer 168 is, optionally, coupled to an input controller 160 in I/O subsystem 106. Accelerometer 168 optionally performs 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 100 optionally includes, in addition to accelerometer(s) 168, 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 100.

[0150] In some embodiments, the software components stored in memory 102 include operating system 126, communication module (or set of instructions) 128, contact/motion module (or set of instructions) 130, graphics module (or set of instructions) 132, text input module (or set of instructions) 134, Global Positioning System (GPS) module (or set of instructions) 135, and applications (or sets of instructions) 136. Furthermore, in some embodiments, memory 102 (FIG. 1A) or 370 (FIG. 3) stores device/global internal state 157, as

shown in FIGS. 1A and 3. Device/global internal state 157 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 112; sensor state, including information obtained from the device's various sensors and input control devices 116; and location information concerning the device's location and/or attitude.

[0151] Operating system 126 (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.

[0152] Communication module 128 facilitates communication with other devices over one or more external ports 124 and also includes various software components for handling data received by RF circuitry 108 and/or external port 124. External port 124 (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.

[0153] Contact/motion module 130 optionally detects contact with touch screen 112 (in conjunction with display controller 156) and other touch-sensitive devices (e.g., a touchpad or physical click wheel). Contact/motion module 130 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 130 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 130 and display controller 156 detect contact on a touchpad.

[0154] In some embodiments, contact/motion module 130 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 100). 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).

[0155] Contact/motion module 130 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.

[0156] Graphics module 132 includes various known software components for rendering and displaying graphics on touch screen 112 or other display, including components for changing the visual impact (e.g., brightness, tran6sparency, 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.

[0157] In some embodiments, graphics module 132 stores data representing graphics to be used. Each graphic is, optionally, assigned a corresponding code. Graphics module 132 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 156.

[0158] Haptic feedback module 133 includes various software components for generating instructions used by tactile output generator(s) 167 to produce tactile outputs at one or more locations on device 100 in response to user interactions with device 100.

[0159] Text input module 134, which is, optionally, a component of graphics module 132, provides soft keyboards for entering text in various applications (e.g., contacts 137, e-mail 140, IM 141, browser 147, and any other application that needs text input).

[0160] GPS module 135 determines the location of the device and provides this information for use in various applications (e.g., to telephone 138 for use in location-based dialing; to camera 143 as picture/video metadata; and to applications that provide location-based services such as weather widgets, local yellow page widgets, and map/navigation widgets).

[0161] Applications 136 optionally include the following modules (or sets of instructions), or a subset or superset thereof:

- Contacts module 137 (sometimes called an address book or contact list);
- Telephone module 138;
- Video conference module 139;
- E-mail client module 140;
- Instant messaging (IM) module 141;

- Workout support module 142;
- Camera module 143 for still and/or video images;
- Image management module 144;
- Video player module;
- Music player module;
- Browser module 147;
- Calendar module 148;
- Widget modules 149, which optionally include one or more of: weather widget 149-1, stocks widget 149-2, calculator widget 149-3, alarm clock widget 149-4, dictionary widget 149-5, and other widgets obtained by the user, as well as user-created widgets 149-6;
- Widget creator module 150 for making user-created widgets 149-6;
- Search module 151;
- Video and music player module 152, which merges video player module and music player module;
- Notes module 153;
- Map module 154; and/or
- Online video module 155.

[0162] Examples of other applications 136 that are, optionally, stored in memory 102 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.

[0163] In conjunction with touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, contacts module 137 are, optionally, used to manage an address book or contact list (e.g., stored in application internal state 192 of contacts module 137 in memory 102 or memory 370), 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 138, video conference module 139, e-mail 140, or IM 141; and so forth.

[0164] In conjunction with RF circuitry 108, audio circuitry 110, speaker 111, microphone 113, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, telephone module 138 are optionally, used to enter a sequence of characters corresponding to a telephone number, access one or more telephone numbers in contacts module 137, 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 optionally uses any of a plurality of communications standards, protocols, and technologies.

[0165] In conjunction with RF circuitry 108, audio circuitry 110, speaker 111, microphone 113, touch screen 112, display controller 156, optical sensor 164, optical sensor controller 158, contact/motion module 130, graphics module 132, text input module 134, contacts module 137, and telephone module 138, video conference module 139 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.

[0166] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, e-mail client module 140 includes executable instructions to create, send, receive, and manage e-mail in response to user instructions. In conjunction with image management module 144, e-mail client module 140 makes it very easy to create and send e-mails with still or video images taken with camera module 143.

[0167] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, the instant messaging module 141 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 optionally 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).

[0168] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, text input module 134, GPS module 135, map module 154, and music player module, workout support module 142 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.

[0169] In conjunction with touch screen 112, display controller 156, optical sensor(s) 164, optical sensor controller 158, contact/motion module 130, graphics module 132, and image management module 144, camera module 143 includes executable instructions to capture still images or video (including a video stream) and store them into memory 102, modify characteristics of a still image or video, or delete a still image or video from memory 102.

[0170] In conjunction with touch screen 112, display controller 156, contact/motion module 130, graphics module 132, text input module 134, and camera module 143, image management module 144 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.

[0171] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, browser module 147 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.

[0172] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, text input module 134, e-mail client module 140, and browser module 147, calendar module 148 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.

[0173] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, text input module 134, and browser module 147, widget modules 149 are mini-applications that are, optionally, downloaded and used by a user (e.g., weather widget 149-1, stocks widget 149-2, calculator widget 149-3, alarm clock widget 149-4, and dictionary widget 149-5) or created by the user (e.g., user-created widget 149-6). 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).

[0174] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, text input module 134, and browser module 147, the widget creator module 150 are, optionally, used by a user to create widgets (e.g., turning a user-specified portion of a web page into a widget).

[0175] In conjunction with touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, search module 151 includes executable instructions to search for text, music, sound, image, video, and/or other files in memory 102 that match one or more search criteria (e.g., one or more user-specified search terms) in accordance with user instructions.

[0176] In conjunction with touch screen 112, display controller 156, contact/motion module 130, graphics module 132, audio circuitry 110, speaker 111, RF circuitry 108, and browser module 147, video and music player module 152 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 112 or on an external, connected display via external port 124). In some embodiments, device 100 optionally includes the functionality of an MP3 player, such as an iPod (trademark of Apple Inc.).

[0177] In conjunction with touch screen 112, display controller 156, contact/motion module 130, graphics module 132, and text input module 134, notes module 153 includes executable instructions to create and manage notes, to-do lists, and the like in accordance with user instructions.

[0178] In conjunction with RF circuitry 108, touch screen 112, display controller 156, contact/motion module 130, graphics module 132, text input module 134, GPS module 135, and browser module 147, map module 154 are, optionally, 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.

[0179] In conjunction with touch screen 112, display controller 156, contact/motion module 130, graphics module 132, audio circuitry 110, speaker 111, RF circuitry 108, text input module 134, e-mail client module 140, and browser module 147, online video module 155 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 124), 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 141, rather than e-mail client module 140, 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.

[0180] 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 are, optionally, combined or otherwise rearranged in various embodiments. For example, video player module is, optionally, combined with music player module into a single module (e.g., video and music player module 152, FIG. 1A). In some embodiments, memory 102 optionally stores a subset of the modules and data structures identified above. Furthermore, memory 102 optionally stores additional modules and data structures not described above.

[0181] In some embodiments, device 100 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 100, the number of physical input control devices (such as push buttons, dials, and the like) on device 100 is, optionally, reduced.

[0182] 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 100 to a main, home, or root menu from any user interface that is displayed on device 100. 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.

[0183] FIG. 1B is a block diagram illustrating exemplary components for event handling in accordance with some embodiments. In some embodiments, memory 102 (FIG. 1A) or 370 (FIG. 3) includes event sorter 170 (e.g., in operating system 126) and a respective application 136-1 (e.g., any of the aforementioned applications 137-151, 155, 380-390).

[0184] Event sorter 170 receives event information and determines the application 136-1 and application view 191 of application 136-1 to which to deliver the event information. Event sorter 170 includes event monitor 171 and event dispatcher module 174. In some embodiments, application 136-1 includes application internal state 192, which indicates the current application view(s) displayed on touch-sensitive display 112 when the application is active or executing. In some embodiments, device/global internal state 157 is used by event sorter 170 to determine which application(s) is (are) currently active, and application internal state 192 is used by event sorter 170 to determine application views 191 to which to deliver event information.

[0185] In some embodiments, application internal state 192 includes additional information, such as one or more of: resume information to be used when application 136-1 resumes execution, user interface state information that indicates information being displayed or that is ready for display by application 136-1, a state queue for enabling the user to go back to a prior state or view of application 136-1, and a redo/undo queue of previous actions taken by the user.

[0186] Event monitor 171 receives event information from peripherals interface 118. Event information includes information about a sub-event (e.g., a user touch on touch-sensitive display 112, as part of a multi-touch gesture). Peripherals interface 118 transmits information it receives from I/O subsystem 106 or a sensor, such as proximity sensor 166, accelerometer(s) 168, and/or microphone 113 (through audio circuitry 110). Information that peripherals interface 118 receives from I/O subsystem 106 includes information from touch-sensitive display 112 or a touch-sensitive surface.

[0187] In some embodiments, event monitor 171 sends requests to the peripherals interface 118 at predetermined intervals. In response, peripherals interface 118 transmits event information. In other embodiments, peripherals interface 118 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).

[0188] In some embodiments, event sorter 170 also includes a hit view determination module 172 and/or an active event recognizer determination module 173.

[0189] Hit view determination module 172 provides software procedures for determining where a sub-event has taken place within one or more views when touch-sensitive display 112 displays more than one view. Views are made up of controls and other elements that a user can see on the display.

[0190] 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 optionally 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 is, optionally, called the hit view, and the set of events that are recognized as proper inputs are, optionally, determined based, at least in part, on the hit view of the initial touch that begins a touch-based gesture.

[0191] Hit view determination module 172 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 172 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 172, 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.

[0192] Active event recognizer determination module 173 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 173 determines that only the hit view should receive a particular sequence of sub-events. In other embodiments, active event recognizer determination module 173 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.

[0193] Event dispatcher module 174 dispatches the event information to an event recognizer (e.g., event recognizer 180). In embodiments including active event recognizer determination module 173, event dispatcher module 174 delivers the event information to an event recognizer determined by active event recognizer determination module 173. In some embodiments, event dispatcher module 174 stores in an event queue the event information, which is retrieved by a respective event receiver 182.

[0194] In some embodiments, operating system 126 includes event sorter 170. Alternatively, application 136-1 includes event sorter 170. In yet other embodiments, event sorter 170 is a stand-alone module, or a part of another module stored in memory 102, such as contact/motion module 130.

[0195] In some embodiments, application 136-1 includes a plurality of event handlers 190 and one or more application views 191, each of which includes instructions for handling touch events that occur within a respective view of the application's user interface. Each application view 191 of the application 136-1 includes one or more event recognizers 180. Typically, a respective application view 191 includes a plurality of event recognizers 180. In other embodiments, one or more of event recognizers 180 are part of a separate module, such as a user interface kit (not shown) or a higher level object from which application 136-1 inherits methods and other properties. In some embodiments, a respective event handler 190 includes one or more of: data updater 176, object updater 177, GUI updater 178, and/or event data 179 received from event sorter 170. Event handler 190 optionally utilizes or calls data updater 176, object updater 177, or GUI updater 178 to update the application internal state 192. Alternatively, one or more of the application views 191 include one or more respective event handlers 190. Also, in some embodiments, one or more of data updater 176, object updater 177, and GUI updater 178 are included in a respective application view 191.

[0196] A respective event recognizer 180 receives event information (e.g., event data 179) from event sorter 170 and identifies an event from the event information. Event recognizer 180 includes event receiver 182 and event comparator 184. In some embodiments, event recognizer 180 also includes at least a subset of: metadata 183, and event delivery instructions 188 (which optionally include sub-event delivery instructions).

[0197] Event receiver 182 receives event information from event sorter 170. 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 optionally also includes 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.

[0198] Event comparator 184 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 184 includes event definitions 186. Event definitions 186 contain definitions of events (e.g., predefined sequences of sub-events), for example, event 1 (187-1), event 2 (187-2), and others. In some embodiments, sub-events in an event (187) include, for example, touch begin, touch end, touch movement, touch cancellation, and multiple touching. In one example, the definition for event 1 (187-1) 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 (187-2) 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 112, and liftoff of the touch (touch end). In some embodiments, the event also includes information for one or more associated event handlers 190.

[0199] In some embodiments, event definition 187 includes a definition of an event for a respective user-interface object. In some embodiments, event comparator 184 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 112, when a touch is detected on touch-sensitive display 112, event comparator 184 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 190, the event comparator uses the result of the hit test to determine which event handler 190 should be activated. For example, event comparator 184 selects an event handler associated with the sub-event and the object triggering the hit test.

[0200] In some embodiments, the definition for a respective event (187) 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.

[0201] When a respective event recognizer 180 determines that the series of sub-events do not match any of the events in event definitions 186, the respective event recognizer 180 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.

[0202] In some embodiments, a respective event recognizer 180 includes metadata 183 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 183 includes configurable properties, flags, and/or lists that indicate how event recognizers interact, or are enabled to interact, with one another. In some embodiments, metadata 183 includes configurable properties, flags, and/or lists that indicate whether sub-events are delivered to varying levels in the view or programmatic hierarchy.

[0203] In some embodiments, a respective event recognizer 180 activates event handler 190 associated with an event when one or more particular sub-events of an event are recognized. In some embodiments, a respective event recognizer 180 delivers event information associated with the event to event handler 190. Activating an event handler 190 is distinct from sending (and deferred sending) sub-events to a respective hit view. In some embodiments, event recognizer 180 throws a flag associated with the recognized event, and event handler 190 associated with the flag catches the flag and performs a predefined process.

[0204] In some embodiments, event delivery instructions 188 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.

[0205] In some embodiments, data updater 176 creates and updates data used in application 136-1. For example, data updater 176 updates the telephone number used in contacts module 137, or stores a video file used in video player module. In some embodiments, object updater 177 creates and updates objects used in application 136-1. For example, object updater 177 creates a new user-interface object or updates the position of a user-interface object. GUI updater 178 updates the GUI. For example, GUI updater 178 prepares display information and sends it to graphics module 132 for display on a touch-sensitive display.

[0206] In some embodiments, event handler(s) 190 includes or has access to data updater 176, object updater 177, and GUI updater 178. In some embodiments, data updater 176, object updater 177, and GUI updater 178 are included in a single module of a respective application 136-1 or application view 191. In other embodiments, they are included in two or more software modules.

[0207] 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 100 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.

[0208] FIG. 2 illustrates a portable multifunction device 100 having a touch screen 112 in accordance with some embodiments. The touch screen optionally displays one or more graphics within user interface (UI) 200. 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 202 (not drawn to scale in the figure) or one or more styluses 203 (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 100. 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.

[0209] Device 100 optionally also include one or more physical buttons, such as "home" or menu button 204. As described previously, menu button 204 is, optionally, used to navigate to any application 136 in a set of applications that are, optionally, executed on device 100. Alternatively, in some embodiments, the menu button is implemented as a soft key in a GUI displayed on touch screen 112.

[0210] In some embodiments, device 100 includes touch screen 112, menu button 204, push button 206 for powering the device on/off and locking the device, volume adjustment button(s) 208, subscriber identity module (SIM) card slot 210, headset jack 212, and docking/charging external port 124. Push button 206 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 100 also accepts verbal input for activation or deactivation of some functions through microphone 113. Device 100 also, optionally, includes one or more contact intensity sensors 165 for detecting intensity of contacts on touch screen 112 and/or one or more tactile output generators 167 for generating tactile outputs for a user of device 100.

[0211] FIG. 3 is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface in accordance with some embodiments. Device 300 need not be portable. In some embodiments, device 300 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 300 typically includes one or more processing units (CPUs) 310, one or more network or other communications interfaces 360, memory 370, and one or more communication buses 320 for interconnecting these components. Communication buses 320 optionally include circuitry (sometimes called a chipset) that interconnects and controls communications between system components. Device 300 includes input/output (I/O) interface 330 comprising display 340, which is typically a touch screen display. I/O interface 330 also optionally includes a keyboard and/or mouse (or other pointing device) 350 and touchpad 355, tactile output generator 357 for generating tactile outputs on device 300 (e.g., similar to tactile output generator(s) 167 described above with reference to FIG. 1A), sensors 359 (e.g., optical, acceleration, proximity, touch-sensitive, and/or contact intensity sensors similar to contact intensity sensor(s) 165 described above with reference to FIG. 1A). Memory 370 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 370 optionally includes one or more storage devices remotely located from CPU(s) 310. In some embodiments, memory 370 stores programs, modules, and data structures analogous to the programs, modules, and data structures stored in memory 102 of portable multifunction device 100 (FIG. 1A), or a subset thereof. Furthermore, memory 370 optionally stores additional programs, modules, and data structures not present in memory 102 of portable multifunction device 100. For example, memory 370 of device 300 optionally stores drawing module 380, presentation module 382, word processing module 384, website creation module 386, disk authoring module 388, and/or spreadsheet module 390, while memory 102 of portable multifunction device 100 (FIG. 1A) optionally does not store these modules.

[0212] Each of the above-identified elements in FIG. 3 is, optionally, 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 are, optionally, combined or otherwise rearranged in various embodiments. In some embodiments, memory 370 optionally

stores a subset of the modules and data structures identified above. Furthermore, memory 370 optionally stores additional modules and data structures not described above.

[0213] Attention is now directed towards embodiments of user interfaces that are, optionally, implemented on, for example, portable multifunction device 100.

[0214] FIG. 4A illustrates an exemplary user interface for a menu of applications on portable multifunction device 100 in accordance with some embodiments. Similar user interfaces are, optionally, implemented on device 300. In some embodiments, user interface 400 includes the following elements, or a subset or superset thereof:

- Signal strength indicator(s) 402 for wireless communication(s), such as cellular and Wi-Fi signals;
- Time 404;
- Bluetooth indicator 405;
- Battery status indicator 406;
- Tray 408 with icons for frequently used applications, such as:
    - Icon 416 for telephone module 138, labeled "Phone," which optionally includes an indicator 414 of the number of missed calls or voicemail messages;
    - Icon 418 for e-mail client module 140, labeled "Mail," which optionally includes an indicator 410 of the number of unread e-mails;
    - Icon 420 for browser module 147, labeled "Browser;" and
    - Icon 422 for video and music player module 152, also referred to as iPod (trademark of Apple Inc.) module 152, labeled "iPod;" and
- Icons for other applications, such as:
    - Icon 424 for IM module 141, labeled "Messages;"
    - Icon 426 for calendar module 148, labeled "Calendar;"

- Icon 428 for image management module 144, labeled "Photos;"
- Icon 430 for camera module 143, labeled "Camera;"
- Icon 432 for online video module 155, labeled "Online Video;"
- Icon 434 for stocks widget 149-2, labeled "Stocks;"
- Icon 436 for map module 154, labeled "Maps;"
- Icon 438 for weather widget 149-1, labeled "Weather;"
- Icon 440 for alarm clock widget 149-4, labeled "Clock;"
- Icon 442 for workout support module 142, labeled "Workout Support;"
- Icon 444 for notes module 153, labeled "Notes;" and
- Icon 446 for a settings application or module, labeled "Settings," which provides access to settings for device 100 and its various applications 136.

[0215] It should be noted that the icon labels illustrated in FIG. 4A are merely exemplary. For example, icon 422 for video and music player module 152 is 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.

[0216] FIG. 4B illustrates an exemplary user interface on a device (e.g., device 300, FIG. 3) with a touch-sensitive surface 451 (e.g., a tablet or touchpad 355, FIG. 3) that is separate from the display 450 (e.g., touch screen display 112). Device 300 also, optionally, includes one or more contact intensity sensors (e.g., one or more of sensors 359) for detecting intensity of contacts on touch-sensitive surface 451 and/or one or more tactile output generators 357 for generating tactile outputs for a user of device 300.

[0217] Although some of the examples that follow will be given with reference to inputs on touch screen display 112 (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. 4B. In some embodiments, the touch-sensitive surface (e.g., 451 in

FIG. 4B) has a primary axis (e.g., 452 in FIG. 4B) that corresponds to a primary axis (e.g., 453 in FIG. 4B) on the display (e.g., 450). In accordance with these embodiments, the device detects contacts (e.g., 460 and 462 in FIG. 4B) with the touch-sensitive surface 451 at locations that correspond to respective locations on the display (e.g., in FIG. 4B, 460 corresponds to 468 and 462 corresponds to 470). In this way, user inputs (e.g., contacts 460 and 462, and movements thereof) detected by the device on the touch-sensitive surface (e.g., 451 in FIG. 4B) are used by the device to manipulate the user interface on the display (e.g., 450 in FIG. 4B) 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.

[0218] 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.

[0219] FIG. 5A illustrates exemplary personal electronic device 500. Device 500 includes body 502. In some embodiments, device 500 can include some or all of the features described with respect to devices 100 and 300 (e.g., FIGS. 1A-4B). In some embodiments, device 500 has touch-sensitive display screen 504, hereafter touch screen 504. Alternatively, or in addition to touch screen 504, device 500 has a display and a touch-sensitive surface. As with devices 100 and 300, in some embodiments, touch screen 504 (or the touch-sensitive surface) optionally includes one or more intensity sensors for detecting intensity of contacts (e.g., touches) being applied. The one or more intensity sensors of touch screen 504 (or the touch-sensitive surface) can provide output data that represents the intensity of touches. The user interface of device 500

can respond to touches based on their intensity, meaning that touches of different intensities can invoke different user interface operations on device 500.

[0220] Exemplary techniques for detecting and processing touch intensity are 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, published as WIPO Publication No. WO/2013/169849, 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, published as WIPO Publication No. WO/2014/105276, each of which is hereby incorporated by reference in their entirety.

[0221] In some embodiments, device 500 has one or more input mechanisms 506 and 508. Input mechanisms 506 and 508, if included, can be physical. Examples of physical input mechanisms include push buttons and rotatable mechanisms. In some embodiments, device 500 has one or more attachment mechanisms. Such attachment mechanisms, if included, can permit attachment of device 500 with, for example, hats, eyewear, earrings, necklaces, shirts, jackets, bracelets, watch straps, chains, trousers, belts, shoes, purses, backpacks, and so forth. These attachment mechanisms permit device 500 to be worn by a user.

[0222] FIG. 5B depicts exemplary personal electronic device 500. In some embodiments, device 500 can include some or all of the components described with respect to FIGS. 1A, 1B, and 3. Device 500 has bus 512 that operatively couples I/O section 514 with one or more computer processors 516 and memory 518. I/O section 514 can be connected to display 504, which can have touch-sensitive component 522 and, optionally, intensity sensor 524 (e.g., contact intensity sensor). In addition, I/O section 514 can be connected with communication unit 530 for receiving application and operating system data, using Wi-Fi, Bluetooth, near field communication (NFC), cellular, and/or other wireless communication techniques. Device 500 can include input mechanisms 506 and/or 508. Input mechanism 506 is, optionally, a rotatable input device or a depressible and rotatable input device, for example. Input mechanism 508 is, optionally, a button, in some examples.

[0223] Input mechanism 508 is, optionally, a microphone, in some examples. Personal electronic device 500 optionally includes various sensors, such as GPS sensor 532, accelerometer 534, directional sensor 540 (e.g., compass), gyroscope 536, motion sensor 538, and/or a combination thereof, all of which can be operatively connected to I/O section 514.

[0224] Memory 518 of personal electronic device 500 can include one or more non-transitory computer-readable storage mediums, for storing computer-executable instructions, which, when executed by one or more computer processors 516, for example, can cause the computer processors to perform the techniques described below, including processes 700, 900, 1100, 1300, 1500, 1700, 1800, 2000, 2200, 2400, 2600, 2800, 3000, and 3200 (FIGS. 7, 9, 11, 13, 15, 17, 18, 20, 22, 24, 26, 28, 30, and 32). A 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. In some examples, the storage medium is a transitory computer-readable storage medium. In some examples, the storage medium is a non-transitory computer-readable storage medium. 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 500 is not limited to the components and configuration of FIG. 5B, but can include other or additional components in multiple configurations.

[0225] As used here, the term "affordance" refers to a user-interactive graphical user interface object that is, optionally, displayed on the display screen of devices 100, 300, and/or 500 (FIGS. 1A, 3, and 5A-5B). For example, an image (e.g., icon), a button, and text (e.g., hyperlink) each optionally constitute an affordance.

[0226] 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 355 in FIG. 3 or touch-sensitive surface 451 in FIG. 4B) 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 112 in FIG. 1A or touch screen 112 in FIG. 4A) 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).

[0227] 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 optionally includes 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.

[0228] FIG. 5C illustrates detecting a plurality of contacts 552A-552E on touch-sensitive display screen 504 with a plurality of intensity sensors 524A-524D. FIG. 5C additionally includes intensity diagrams that show the current intensity measurements of the intensity sensors 524A-524D relative to units of intensity. In this example, the intensity measurements of intensity sensors 524A and 524D are each 9 units of intensity, and the intensity measurements of intensity sensors 524B and 524C are each 7 units of intensity. In some implementations, an aggregate intensity is the sum of the intensity measurements of the plurality of intensity sensors 524A-524D, which in this example is 32 intensity units. In some embodiments, each contact is assigned a respective intensity that is a portion of the aggregate intensity. FIG. 5D illustrates assigning the aggregate intensity to contacts 552A-552E based on their distance from the center of force 554. In this example, each of contacts 552A, 552B, and 552E are assigned an intensity of contact of 8 intensity units of the aggregate intensity, and each of contacts 552C and 552D are assigned an intensity of contact of 4 intensity units of the aggregate intensity. More generally, in some implementations, each contact j is assigned a respective intensity Ij that is a portion of the

aggregate intensity, A, in accordance with a predefined mathematical function, $Ij = A \cdot (Dj/\Sigma Di)$, where Dj is the distance of the respective contact j to the center of force, and ΣDi is the sum of the distances of all the respective contacts (e.g., i=1 to last) to the center of force. The operations described with reference to FIGS. 5C-5D can be performed using an electronic device similar or identical to device 100, 300, or 500. In some embodiments, a characteristic intensity of a contact is based on one or more intensities of the contact. In some embodiments, the intensity sensors are used to determine a single characteristic intensity (e.g., a single characteristic intensity of a single contact). It should be noted that the intensity diagrams are not part of a displayed user interface, but are included in FIGS. 5C-5D to aid the reader.

[0229] In some embodiments, a portion of a gesture is identified for purposes of determining a characteristic intensity. For example, a touch-sensitive surface optionally receives 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 is, optionally, 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 is, optionally, 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.

[0230] The intensity of a contact on the touch-sensitive surface is, optionally, 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.

[0231] 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.

[0232] 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).

[0233] FIGS. 5E-5H illustrate detection of a gesture that includes a press input that corresponds to an increase in intensity of a contact 562 from an intensity below a light press intensity threshold (e.g., "IT_L") in FIG. 5E, to an intensity above a deep press intensity threshold (e.g., "IT_D") in FIG. 5H. The gesture performed with contact 562 is detected on touch-sensitive surface 560 while cursor 576 is displayed over application icon 572B corresponding to App 2, on a displayed user interface 570 that includes application icons 572A-572D displayed in predefined region 574. In some embodiments, the gesture is detected on touch-sensitive display 504. The intensity sensors detect the intensity of contacts on touch-sensitive surface 560. The device determines that the intensity of contact 562 peaked above the deep press intensity threshold (e.g., "IT_D"). Contact 562 is maintained on touch-sensitive surface 560. In response to the detection of the gesture, and in accordance with contact 562 having an intensity that goes above the deep press intensity threshold (e.g., "IT_D") during the gesture, reduced-scale representations 578A-578C (e.g., thumbnails) of recently opened documents for App 2 are displayed, as shown in FIGS. 5F-5H. In some embodiments, the intensity, which is compared to the one or more intensity thresholds, is the characteristic intensity of a contact. It should be noted that the intensity diagram for contact 562 is not part of a displayed user interface, but is included in FIGS. 5E-5H to aid the reader.

[0234] In some embodiments, the display of representations 578A-578C includes an animation. For example, representation 578A is initially displayed in proximity of application icon 572B, as shown in FIG. 5F. As the animation proceeds, representation 578A moves upward and representation 578B is displayed in proximity of application icon 572B, as shown in FIG. 5G. Then, representations 578A moves upward, 578B moves upward toward representation 578A, and representation 578C is displayed in proximity of application icon 572B, as shown in FIG. 5H. Representations 578A-578C form an array above icon 572B. In some embodiments, the animation progresses in accordance with an intensity of contact 562, as shown in FIGS. 5F-5G, where the representations 578A-578C appear and move upwards as the intensity of contact 562 increases toward the deep press intensity threshold (e.g., "IT_D"). In some embodiments, the

intensity, on which the progress of the animation is based, is the characteristic intensity of the contact. The operations described with reference to FIGS. 5E-5H can be performed using an electronic device similar or identical to device 100, 300, or 500.

[0235] 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).

[0236] 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.

[0237] FIG. 5I illustrates exemplary electronic device 580. Device 580 includes body 580A. In some embodiments, device 580 can include some or all of the features described with respect to devices 100, 300, and 500 (e.g., FIGS. 1A-5B). In some embodiments, device 580 has one or more speakers 580B (concealed in body 580A), one or more microphones 580C, one or more touch-sensitive surfaces 580D, and one or more displays 580E. Alternatively, or in addition to a display and touch-sensitive surface 580D, the device has a touch-sensitive display (also referred to as a touchscreen). As with devices 100, 300, and 500, in some embodiments, touch-sensitive surface 580D (or the touch screen) optionally includes one or more intensity sensors for detecting intensity of contacts (e.g., touches) being applied. The one or more intensity sensors of touch-sensitive surface 580D (or the touchscreen) can provide output data that represents the intensity of touches. The user interface of device 580 can respond to touches based on their intensity, meaning that touches of different intensities can invoke different user interface operations on device 580. In some embodiments, the one or more displays 580E are one or more light-emitting diodes (LEDs). For example, a display can be a single LED, an LED cluster (e.g., a red, a green, and a blue LED), a plurality of discrete LEDs, a plurality of discrete LED clusters, or other arrangement of one or more LEDs. For example, the display 580E can be an array of nine discrete LED clusters arranged in a circular shape (e.g., a ring). In some examples, the one or more displays are comprised of one or more of another type of light-emitting elements.

[0238] FIG. 5J depicts exemplary personal electronic device 580. In some embodiments, device 580 can include some or all of the components described with respect to FIGS. 1A, 1B, 3, and 5A-5B. Device 580 has bus 592 that operatively couples I/O section 594 with one or more computer processors 596 and memory 598. I/O section 594 can be connected to display 582, which can have touch-sensitive component 584 and, optionally, intensity sensor 585 (e.g., contact intensity sensor). In some embodiments, touch-sensitive component 584 is a separate component than display 582. In addition, I/O section 594 can be connected with communication unit 590 for receiving application and operating system data, using Wi-Fi, Bluetooth, near field communication (NFC), cellular, and/or other wireless communication techniques. Device 580 can include input mechanisms 588. Input mechanism 588 is, optionally, a button, in some

examples. Input mechanism 588 is, optionally, a microphone, in some examples. Input mechanism 588 is, optionally, a plurality of microphones (e.g., a microphone array).

[0239] Electronic device 580 includes speaker 586 for outputting audio. Device 580 can include audio circuitry (e.g., in I/O section 594) that receives audio data, converts the audio data to an electrical signal, and transmits the electrical signal to speaker 586. Speaker 586 converts the electrical signal to human-audible sound waves. The audio circuitry (e.g., in I/O section 594) also receives electrical signals converted by a microphone (e.g., input mechanism 588) from sound waves. The audio circuitry (e.g., in I/O section 594) converts the electrical signal to audio data. Audio data is, optionally, retrieved from and/or transmitted to memory 598 and/or RF circuitry (e.g., in communication unit 590) by I/O section 594.

[0240] Memory 598 of personal electronic device 580 can include one or more non-transitory computer-readable storage mediums, for storing computer-executable instructions, which, when executed by one or more computer processors 596, for example, can cause the computer processors to perform the techniques described below, including processes 700, 900, 1100, 1300, 1500, 1700, 1800, 2000, 2200, 2400, 2600, 2800, 3000, and 3200 (FIGS. 7, 9, 11, 13, 15, 17, 18, 20, 22, 24, 26, 28, 30, and 32). A 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. In some examples, the storage medium is a transitory computer-readable storage medium. In some examples, the storage medium is a non-transitory computer-readable storage medium. 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 580 is not limited to the components and configuration of FIG. 5J, but can include other or additional components in multiple configurations.

[0241] As used herein, an "installed application" refers to a software application that has been downloaded onto an electronic device (e.g., devices 100, 300, and/or 500) and is ready to be launched (e.g., become opened) on the device. In some embodiments, a downloaded

application becomes an installed application by way of an installation program that extracts program portions from a downloaded package and integrates the extracted portions with the operating system of the computer system.

[0242] As used herein, the terms "open application" or "executing application" refer to a software application with retained state information (e.g., as part of device/global internal state 157 and/or application internal state 192). An open or executing application is, optionally, any one of the following types of applications:

- an active application, which is currently displayed on a display screen of the device that the application is being used on;
- a background application (or background processes), which is not currently displayed, but one or more processes for the application are being processed by one or more processors; and
- a suspended or hibernated application, which is not running, but has state information that is stored in memory (volatile and non-volatile, respectively) and that can be used to resume execution of the application.

[0243] As used herein, the term "closed application" refers to software applications without retained state information (e.g., state information for closed applications is not stored in a memory of the device). Accordingly, closing an application includes stopping and/or removing application processes for the application and removing state information for the application from the memory of the device. Generally, opening a second application while in a first application does not close the first application. When the second application is displayed and the first application ceases to be displayed, the first application becomes a background application.

[0244] Attention is now directed towards embodiments of user interfaces ("UI") and associated processes that are implemented on an electronic device, such as portable multifunction device 100, device 300, or device 500.

[0245] FIGS. 6A-6AR illustrate exemplary user interfaces for accessing media playback status of a plurality of available playback devices, in accordance with some embodiments. The

user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 7.

[0246] FIGS. 6A-6B illustrate exemplary user interfaces for managing and controlling a group of devices. In some embodiments, the group of devices includes devices that are placed throughout a user's home to perform various functions. Such functions can include controlling smart devices (e.g., light bulbs, thermostats, media devices), configuring home automation, and controlling home entertainment.

[0247] FIG. 6A depicts home summary screen 601 (on display 602 of device 600), which represents an exemplary interface for a home control application (also referred to as a "device management application") executing on device 600. In some embodiments, device 600 includes one or more features of device 100, 300, or 500. An example of a home control application is "Home" (by Apple Inc., of Cupertino, California, USA). The home summary screen 601 includes scenes (e.g., pre-configured collection of settings for one more home devices) that have been stored, as well as an indication of devices in the user's home (also referred to as "home devices" or the like). The home devices are represented by indicators 601A-601C. Indicator 601A corresponds to the device named "Smart Speaker 1" in the room "Living Room". Indicator 601B corresponds to the device named "Smart Speaker 2" in the room "Kitchen". Indicator 601C corresponds to the device named "Smart Bulb" in the room "Bedroom". Exemplary interfaces and techniques for a home control application described in U.S. Patent Application 62/349,057, "User Interface for Managing Controllable External Devices," filed June 12, 2016, which is hereby incorporated by reference in its entirety.

[0248] FIG. 6A also depicts exemplary home configuration screen 603, which illustrates an exemplary interface for managing configuration settings associated with the home devices represented in home summary screen 601. In some examples, device 600 displays home configuration screen 603 in response to user input selection of affordance 601D at home summary screen 601. At home configuration screen 603, a user can edit the name of the home ("Home" in this example), edit the members of a home (e.g., Users 1, 2, and 3 in this example), include notes to other members, among other functions. As can be seen, the "Home" in this example includes three members, User 1 (represented by indicator 603A), User 2 (represented by

indicator 603B), and User 3 (represented by indicator 603C). In some examples, Users 1-3 represent users that live in the location that physically includes the home devices, collectively referred to as the "Home". Thus, Users 1-3 are members of the home. Being a member of a home (or other location), in some examples, provides the users with permission to control devices associated with the home. The users can control the home devices using one or more personal devices associated with at least one of the users. In some examples, a device is associated with a user when it is logged into an account associated with (e.g., owned by) the user. For example, devices that are logged into one or more user accounts associated with exemplary Users 1-3 described above are associated with at least one user, and specifically, at least one user of the Home (e.g., home devices).

[0249] FIG. 6B depicts exemplary user interfaces of a portable electronic device 600. Device home screen 605, displayed on display 602, is an exemplary default home screen of device 600. Device home screen 605 includes affordances associated with applications stored on the device. For example, affordance 605A corresponds to a music application. Selection of affordance 605A, in some examples, causes the device to display a user interface of the corresponding music application. FIG. 6B also depicts exemplary user input 604, received by device 600 while displaying screen 603. In this example, user input 604 is a swipe gesture (e.g., touch input followed by a drag) on a touch-sensitive display 602 of device 600. In the example depicted in FIG. 6B, the user input 604 is a swipe gesture from an edge of the touch-sensitive display 602. In some embodiments, user input 604 is an arcuate swipe gesture. For example, an arcuate swipe gesture is a gesture than includes both a horizontal component (e.g., x-direction) and a vertical component (e.g., y-direction) that satisfy arcuate criteria. For instance, user input 604 can be a touch and drag on a touch-sensitive surface that traces the form of a quarter-circle. In response to input 604, device 600 displays a Control Center interface (e.g., FIG. 6C and 6D).

[0250] FIG. 6C illustrates an exemplary control center interface. Control center interface 606 includes various affordances for controlling various functions and/or components of device 600. In the example depicted, control center interface 606 includes exemplary illustrations of: a media control interface 606A, a volume control interface 606C, a screen brightness control interface 606D, a camera affordance 606E, a calculator affordance 606F, a timer affordance 606G, a flashlight affordance 606H, a wireless media streaming affordance 606I, a screen

rotation lock affordance 606J, a do not disturb affordance 606K, wireless antenna control interface 606L. Exemplary media control interface 606A includes media playback controls 606M (e.g., previous track, pause/play, next track) and a device selection affordance 606B. Exemplary wireless antenna control affordances 606L includes an affordance 606N for toggling (e.g., turning on and turning off) a Wi-Fi function of device 600.

[0251] FIG. 6D illustrates another exemplary Control Center interface 606 when device 600 is playing back media through an external device. As shown in FIG. 6D, the device selection affordance 606B includes an indication that the media that is currently being played back (e.g., the song track "Same Thing" by the artist "Jambug") is being played back though an external device (or devices) named "AirPods". Thus, the media control interface 606A indicates the media currently being played (otherwise referred to as being played back, currently playing back, or the like) as well as the device on which the audio output is being routed. In this case, the audio output is generated by wireless personal earbuds ("AirPods" by Apple Inc. of Cupertino, California USA). FIG. 6D also shows volume control affordance 606C including an indicator 606M representative of the "AirPods" device, indicating that the volume control affordance 606C can be used to adjust the volume of the audio output of the external device "AirPods".

[0252] FIG. 6E illustrates the exemplary Control Center interface, as shown in FIG. 6C, but with a user input 608A on device selection affordance 606B. In some embodiments, device 600 displays a multi-device interface (e.g., as shown in FIG. 6J) in response to receiving user input 608A. For example, user input 608A is a touch on the device selection affordance (e.g., also referred to as a "tap" or a "press").

[0253] FIG. 6F illustrates the exemplary Control Center interface, as shown in FIG. 6C, but with a user input 608B at a location on the media control interface 606A. In some embodiments, device 600 displays a multi-device interface (e.g., as shown in FIG. 6J) in response to receiving user input 608B. In some examples, user input 608B is a deep press (e.g., a touch having a characteristic intensity that meets an intensity-dependent criteria). In some examples, an intensity-dependent criteria is whether a characteristic intensity of a touch exceeds a threshold intensity. In some examples, user input 608B is a tap and hold (e.g., a touch input that is continuously detected on a touch-sensitive surface of device 600 for greater than a threshold

length of time). In some examples, the user input 608B (e.g., whether a deep press, or a tap and hold input) can be received at any location (e.g., completely or partially within) on media control interface 606A. For example, in response to receiving the user input 606B on a playback control affordance (e.g., 606M), the device 600 performs an operation different than it would had the user input been a tap (e.g., the device 600 performs a second operation, rather than a playback control operation such as pausing playback).

[0254] FIG. 6G illustrates an exemplary expanded media control interface for a device. In some embodiments, device 600 displays expanded media control interface 609 of FIG. 6G in response to receiving user input 608B. As shown, the expanded media control interface 609 includes an indicator 609A representing media that is currently being played back on the device 600. In this example, the device 600 is named "iPhone" and is currently playing the media track "Same Thing" by the artist "Jambug". Indicator 609A includes a graphic to the left of the track name, representing an album cover associated with the media track. The graphic included in the indicator 609A, in some examples, can be any graphic that is associated with a media track or that can otherwise provide a visual representation of the currently playing media. For example, other appropriate graphic indicators include, but are not limited to, an artist photo, an album cover, a movie cover, a movie poster, a stylized depiction of text, or the like. The exemplary expanded media control interface 609 includes a media progress indicator 609B (e.g., depicting temporal progress through the media track (time elapsed), and the amount of time until the end of the current media (time remaining)). Exemplary expanded media control interface 609 also includes playback control affordances 609C (e.g., that function similar or the same to those described above), volume control affordance 609D (e.g., that function similar or the same to those described above), and a device selection affordance 609E.

[0255] FIGS. 6H-6N illustrate exemplary multi-device interfaces in accordance with some embodiments. FIG. 6H depicts an exemplary multi-device interface 610 that includes affordances 610B-610D for the current device 600. In some embodiments, device 600 displays multi-device interface 610 in response to receiving user input 608A. In some embodiments, device 600 displays multi-device interface 610 in response to receiving user input 608B.

[0256] As shown in FIG. 6H, multi-device interface 610 includes, as described above regarding expanded media control interface 609, an indicator 610A representing media (that is currently being played back on the device 600, which is named "iPhone"), a media progress indicator 610B (e.g., depicting temporal progress through the media track (time elapsed), and the amount of time until the end of the current media (time remaining)). Exemplary multi-device interface 610 also includes playback control affordances 610C (e.g., that function similar or the same to those described above), volume control affordance 610D (e.g., that function similar or the same to those described above), and a device selection affordance 610E. In contrast to the expanded media control interface 609 of FIG. 6G, multi-device interface 610 provides the user with information on the media playback status of multiple devices, as well as the ability to affect the same.

[0257] As shown in FIG. 6H, multi-device interface 610 also includes indicators associated with other playback devices. Indicator 610F corresponds to the one or more devices that are named (or in a set of devices named) "Living Room". In some embodiments, a set of devices can include one or more devices. The set of devices can be assigned the name "Living Room" based on their location within the user's home. For example, when configuring (e.g., using device 600 and a home control application) the one or more devices in the set named Living Room (e.g., before first use), the name "Living Room" is assigned to the one or more devices in the set. Thus, when selecting or viewing the playback status of devices throughout the home, the user is provided an intuitive identification of the set of devices by location. In some embodiments, the name of a device or a device set is a default name. For example, if a device is not configured by a user to have a specific name, the device can simply be named "Device 1" or "Device 1 – Living Room" if the location is associated (e.g., by a user or automatically). In some embodiments, the name of a device or a device set is a user-configured name. For example, the user selects a name from a predefined set of names. In some examples, the name displayed for a device (e.g., in a multi-device interface) is the location of the device. For example, a device can be named (e.g., in a configuration setting or a home control application) "Tim's Speaker" but appear as "Tim's Room" based on a location associated with the device. In some embodiments, a set of devices is a custom name. For example, the user can assign personalized names to device sets, such as "Tim's Room". In some embodiments, a name other than a location name is used. For example, the user can give a set of devices an arbitrary name.

[0258] In this example, the set of devices named Living Room and represented by indicator 610F are currently playing back the media item titled "Heart Wave" by the artist "MKTS", which are indicated inside of indicator 610F. In some embodiments, a media item is one or more of a song, a video, a podcast, an audio stream, a video stream, an audio file, a video file, an e-book (e.g., also referred to as an electronic book), a text document, a webpage, a phone call, a video call, or the like. Indicator 610G corresponds to the set of devices named or included in the set named "Bedroom", which are currently playing back the song titled "Vibes" by artist "John Appleseed". In some embodiments, an indicator for a set of devices (e.g., indicator 610F and 610G) is displayed (e.g., in a multi-device interface) even when the corresponding set of devices is not currently playing back media. In some embodiments, an indicator for a set of devices (e.g., indicator 610F and 610G) is displayed (e.g., in a multi-device interface) even when fewer than all devices in the corresponding set of devices are currently playing back media. For example, if the set of Living Room devices includes three devices (e.g., three smart speaker devices, such as device 580), and only one of the devices is currently playing a media item (e.g., "Heart Wave" by "MKTS"), the corresponding indicator in the multi-device interface still includes an identification of currently playing back media (e.g., "Heart Wave" by "MKTS" is displayed inside of indicator 610F for the device set "Living Room").

[0259] In some embodiments, if individual devices of the set of devices are concurrently playing back different media items, the multi-device interface displays an indication that multiple media items are playing on the devices of a set of devices. For example, the multi-device interface can include a single indicator for the set (e.g., 610F) and an identification that multiple media items are being played back on those devices (e.g., "1 movie, 1 song"). For example, the multi-device interface can include two indicators (e.g., 610F) for the set and an identification of each respective media item being played back (e.g., an indicator like 610F for each media item, both identified as "Living Room", or identified as "Living Room – Device 1" and "Living Room – Device 2").

[0260] A set of devices can also be referred to as a set of "available playback devices". In some embodiments, available playback devices include one or more (e.g., remote, external) devices on which a current device (e.g., device 600) can control or affect playback (e.g., using a multi-device interface such as 610). For example, as shown in FIG. 6H, if the set of devices

named Living Room includes three external devices (e.g., 580 of FIG. 5I) that are connected to device 600 via a communication link (e.g., Bluetooth, Wi-Fi, or the like), and can be controlled by instructions issued by device 600 (e.g., representing user input), the devices in the set named Living Room are "available playback devices". In some embodiments, affecting or controlling playback on an available playback device is subject to conditions or restrictions. For example, device 600 can see the devices in the Living Room represented by indicator 610F, but modifying the playback status of those devices can be subject to permission (e.g., if the media being played back on the Living Room devices was initiated by another device or user). In some embodiments, available playback devices include the current device. For example, in the example of the Living Room including three devices, a plurality of available devices includes device 600 as well as the three living room devices, for a total of at least four devices. Thus, multi-device interface 610 provides a convenient interface that provides an indication of what is currently playing on the current device, as well as on other devices.

[0261] In some embodiments, one or more devices automatically detect another device. In some embodiments, the devices then form a network between themselves. For example, two devices (e.g., 580) can detect each other after booting up for the first time (e.g., straight out of the box, without configuration by a user), establish a connection link between themselves (e.g., and between device 600), and appear in a multi-device interface (e.g., with a name such as "Speaker 1", and/or a location such as "No Location"). In some embodiments, the device (e.g., 580) is configured to part of a user's home (e.g., the home devices). One of skill in the art will appreciate that a variety of techniques can be used to interconnect and/or transmit information between devices (e.g., 580, 600), which are all intended to be within the scope of this disclosure, and thus additional details are not included here.

[0262] In FIG. 6I, device 600 receives user input 612 representing selection of device selection affordance 610E. For example, user input 612 can be a tap or touch on a location associated with the displayed device selection affordance 610E.

[0263] FIG. 6J illustrates an exemplary multi-device user interface, in accordance with some embodiments. In some embodiments, device 600 displays multi-device interface 610 in response to receiving user input 608B. In some embodiments, device 600 displays multi-device interface

610 in response to receiving user input 612. In some examples, a multi-device interface is displayed in response to selection of a device selection affordance (e.g., 610E of FIG. 6I) in any media application.

[0264] As shown in FIG. 6J, multi-device interface 610 includes individual indicators 610H-610O for each device of the available playback devices. In the depicted example, the individual indicators are visually associated with the indicator 610A (which identifies the media item "Same Thing" by "Jambug", as well as the devices currently playing back the media, in this case only the device named "iPhone"). Indicator 610H identifies the current device (named "iPhone") and includes the selection indicator 610I. In this example, selection indicator 610I includes a checkmark, which indicates that the device 610H is currently playing back the media of indicator 610A. Indicator 610J identifies the set of devices named "Bedroom" and includes indicator 610K, which does not include a checkmark (e.g., the Bedroom set of devices is not currently playing back the media track "Same Thing" by "Jambug"). Indicator 610J also includes an indication that the Bedroom set of devices are currently playing back other media. For example, in FIG. 6J, the indicator 610J includes the title of the track "Vibes". The visual indication communicates to the user that the Bedroom devices set is currently busy. If the user wishes to add the Bedroom device set to the current playback of "Same Thing" by "Jambug" (e.g., by selecting indicator 610K), doing so would interrupt media that is currently playing in the Bedroom. Indicator 610L identifies the set of devices named "Kitchen" and includes indicator 610M, which does not include a checkmark (e.g., the Kitchen set of devices is not currently playing back the media track "Same Thing" by "Jambug"). Indicator 610N identifies the set of devices named "Living Room" and includes indicator 610O, which does not include a checkmark (e.g., the Living Room set of devices is not currently playing back the media track "Same Thing" by "Jambug"). Indicator 610N also includes an indication that the Living Room set of devices are currently playing back other media. For example, in FIG. 6J, the indicator 610J includes the title of the track "Vibes".

[0265] As shown in FIG. 6J, the individual indicators are visually associated with the indicator 610A. As can be seen, the individual indicators 610H-610O are included within the box outline of indicator 610A. In some examples, this provides a visual indication that the multi-device interface is currently "targeted" or "focused" on the "media session" identified by

indicator 610A. In some embodiments, a media session includes a device and a media item. In the example shown in FIG. 6J, indicator 610A represents a media session that includes the media item "Same Thing" by "Jambug", and the device "iPhone" (e.g., device 600). In some embodiments, a media session includes a plurality of devices. In some embodiments, a media session includes a plurality of media items. For example, a media session can include playback of a playlist or other collection of multiple media items. Visually associating the indicators for individual devices with an indicator for a media session, visually indicates that selecting an individual device will add a device to that media session. This visual indication is useful, for example, when there are multiple concurrent media sessions within the user's home. For example, indicators 610F and 610G represent, respectively, other media sessions (e.g., that represent playback of different media on different devices).

[0266] As shown in FIG. 6J, the indicators for individual devices have replaced the media playback control affordances (e.g., 610B-610D). In some embodiments, indicators for individual devices and media playback control affordances are concurrently displayed.

[0267] In some embodiments, an indicator associated with a set of devices or an individual device includes an identification of a user who initiated playback of media on, respectively, the set of devices or the individual device. For example, with reference to FIG. 6J, when media session in the Bedroom was initiated by a user (e.g., Steve), other than the user (e.g., Tim) that is currently associated with device 600, one or more of indicator 610G or indicator 610J can identify the user who initiated media playback in the Bedroom—for example, "Steve is playing Vibes by John Appleseed".

[0268] The multi-device interface 610 of FIG. 6J also includes affordance 610R. For example, selection of affordance 610R can cause the indicators for individual devices to cease being displayed (and, for example, the media playback control affordances are again displayed). In some examples, if the device 600 received selection of one or more individual indicators 610K, 610M, or 610O (e.g., user input that causes device 600 to display checkmarks inside of the indicators), selection of affordance 610R causes the device 600, in response, to initiate media playback on the corresponding selected devices. In some embodiments, device 600 initiates playback in response to the selection of one or more individual indicators 610K, 610M, or 610O.

[0269] FIG. 6K illustrates device 600 receiving selection of an individual indicator for a device of the available playback devices. User input 614 is received on indicator 610, associated with the Kitchen set of devices (represented by 610L). In some examples, user input 614 is a tap or touch on a location associated with indicator 610M. In some examples, user input 614 is a deep press on a location associated with indicator 610M.

[0270] FIG. 6L illustrates an exemplary multi-device interface 610 after receiving selection (e.g., user input 614) of an individual indicator for a device of the available playback devices. As can be seen in FIG. 6L, the indicator 610A now includes an indication ("Kitchen + iPhone") that the current media session includes the Kitchen set of devices and the device named iPhone (e.g., device 600). Further, indicator 610L has been updated to include an individual volume control affordance, and indicator 610M now includes a checkmark. In some examples, the Kitchen device set is added to the current media session of indicator 610A in response to user input selection of their associated indicator 610M. In some examples, an additional user input (e.g., user input 615) on affordance 610R causes the Kitchen devices to be added to the current media session of indicator 610A.

[0271] FIG. 6M-6N illustrate switching between media sessions using an exemplary multi-device interface when. In FIG. 6M, multi-device interface 610 is shown, with a first media session associated with indicator 610A (e.g., playback of "Same Thing" by artists "Jambug" on the device "iPhone"). As shown in FIG. 6M, device 600 receives user input 616 associated with indicator 610F, which is associated with a second media session (e.g., playback of "Heart Wave" by artists "MKTS" on the "Living Room" set of devices). In some examples, user input 616 is a tap or touch at a location associated with indicator 610F. In some examples, user input 616 is a deep press at a location associated with indicator 610F.

[0272] In some embodiments, in response to user input 616, the device 600 updates the display of the multi-device interface to "focus" (also referred to as "target") the second media session. For example, FIG. 6N depicts indicator 610A, associated with the first media session, as minimized (e.g., with no visually associated indicators for individual available playback devices or media playback controls). However, indicator 610F has been expanded to include individual indicators for each device of the plurality of available playback devices. Alternatively, or in

addition, an expanded indicator for a media session (e.g., indicator 610F of FIG. 6N) can include media playback control affordances. As shown in FIG. 6N, indicators 610H, 610J, 610L, and 610N are now visually associated with indicator 610F for the second media session on the Living Room set of devices. Notably, indicators 610M and 610O now include checkmarks, indicating that they are part of the currently-selected media session (e.g., associated with the expanded indicator 610F).

[0273] Switching between media sessions can be referred to as "retargeting", "targeting", "refocusing", "focusing", or the like. Thus, the selection of indicator 610F can be referred to as "focusing" the current device 600 or "focusing" the multi-device interface 610 on a set of devices or a media session.

[0274] As shown above with respect to FIGS. 6H to 6N, a multi-device interface can be used to view a media playback status of a plurality of available playback devices. Additionally, a multi-device interface can be used to view what each device in a set of available playback devices is currently playing back. Further, individual devices can be added or removed from a media session using the multi-device interface. In addition to the examples shown in FIGS. 6H to 6N, other exemplary techniques for invoking and interacting with the multi-device interface can be used, as described below.

[0275] FIGS. 6O-6Y illustrate exemplary techniques for interacting with a multi-device user interface, in accordance with some embodiments. FIGS. 6O-6R illustrate an exemplary technique for initiating playback directly onto one or more dynamically-selected external devices. FIG. 6O illustrates an exemplary browsing interface. Browsing interface 620 depicts an interface for browsing media items. For example, a user can use a browsing interface to find and select one or more media items for playback. Exemplary browsing interfaces include interfaces for browsing a music library, a video library, media content available for streaming, or the like. In this example, the browsing interface 620 is a music browsing interface. Indicator 620A shows that the device is displaying a list of media items (e.g., a playlist) titled "New Music". Media item 620B (also referred to as song 620B) is a song titled "Monday Morning" by artist "The Fomoers".

[0276] In FIG. 6P, device 600 receives user input 622 on song 620B. In some examples, user input representing selection of a media item (e.g., a tap on a song) causes playback of the song to initiate on the current device. For example, if device 600 is currently only "targeting" itself, then a tap on a media item causes playback to initiate on device 600. For further example, if device 600 is currently targeting itself and the device set "Living Room", then a tap on a media item causes playback to initiate on device 600 and the set of devices named "Living Room". In some examples, a deep press user input or a tap and hold user input on a media item causes device 600 to perform a secondary operation (e.g., other than immediately initiate playback of the corresponding media item on the targeted device or devices). In some examples, the secondary operation allows, before playback of the corresponding media is initiated, selection of one or more devices that the user would like to initiate media playback on. In this way, the user can begin playback directly on the desired set of devices, without needing to first initiate playback on a current device (e.g., device 600) and then transfer playback to one or more external devices (e.g., through a cumbersome arrangement of interfaces). One of the drawbacks of first initiating playback on the current device (e.g., a personal device) and then transferring playback to other devices (e.g., a home theater system) is that by the time media playback is initiated on the desired set of devices, the media item has already begun playing on the device—if playing back for a room full of listeners, such a scenario is undesirable. Even if playback is restarted (manually or automatically) in response to or after transferring playback to the desired set of devices, the original device that initiated playback will have wasted device resources by starting the playback of the song when such operation was not desired by the user. In some examples, a tap input invokes the primary operation, and a deep press or a tap and hold input invokes the secondary operation. One of skill in the art would recognize that any two user inputs, recognizable by a device as different types of user input, can be used to invoke respective primary and secondary operations.

[0277] FIG. 6Q illustrates an exemplary media item menu. For example, media item menu 624 is displayed in response to user input 622, which is a deep press or a tap and hold input (e.g., invoking a secondary operation on a media item). Menu 624 includes a list of possible actions that can be selected for the media item 620B (the song titled "Monday Morning" by "The Fomoers"). Media item menu 624 includes play to speaker affordance 624A. In FIG. 6R, device 600 receives user input 626 on the play to speaker affordance 624A. For example, user input

626 is a tap. In some examples, in response to receiving the user input 626 representing selection of the play to speaker affordance 624A, device 600 displays a multi-device interface (e.g., similar to multi-device interface 610 of FIG. 6J) for selecting one or more desired devices to initiate playback on. For example, a multi-device interface can be displayed that includes an indicator (e.g., 610A) for a media session that includes the media item "Monday Morning" by "The Fomoers" and no selected devices (e.g., no devices with indicators that include checkmarks). In this example, after selection of the one or more desired devices (e.g., to add them to the media session), playback is then initiated on the desired devices (e.g., device 600 initiates playback on itself and/or transmits and instruction to other devices in the set of desired devices). In this example, a user input (e.g., 622) associated with a media item is used to invoke a multi-device interface (e.g., 610) so that one or more external devices can be selected before device 600 initiates playback of the media. Thus, device 600 initiates playback of the media item (e.g., "Monday Morning" as shown in FIG. 6P) from the beginning (e.g., from the start of the song) on the selected one or more external devices, rather than first initiating playback on device 600 and then subsequently transferring playback (e.g., during playback).

[0278] FIGS. 6S-6Y illustrate exemplary techniques for providing an indication of a device's focus. FIG. 6S illustrates an exemplary current media interface. Current media interface 630 (also referred to as a "now playing interface") identifies a media item that is currently being played back. Current media interface 630 includes album art 630A, media identifier 630B (song title "Same Thing"), media identifier 630C (artist "Jambug"), media identifier 630D (album "Fly Away"), and device selection affordance 630E. As shown in FIG. 6S, the device selection affordance 630E does not include an identifier for another device. Thus, in this example, the media item identified in the current media interface 630A is currently playing back on the current device, device 600. In some examples, the media identified by the currently media interface is currently being played back on other devices (e.g., in addition to the current device, or a set that excludes the current device).

[0279] At FIG. 6T, device 600 receives user input 632 on device selection affordance 630E. In response to receiving user input 632, at FIG. 6U the device displays multi-device interface 610. As can be seen in FIG. 6U, a first region 611A (e.g., shown bordered by a first dotted line pattern, for illustration purposes) of the display includes the multi-device interface 610, and a

second region 611B (e.g., shown bordered by a second dotted line pattern, for illustration purposes) of the display includes a visible indication of the content displayed prior to displaying the multi-device interface. In this case, the content displayed prior to displaying the multi-device interface is the current media interface 630. For example, the album art 630A is visible through the multi-device interface. In this example, the first region is visually overlaid on the second region.

[0280] At FIG. 6V, device 600 receives a user input 634 on indicator 610F. In some examples, user input 634 is a tap or touch at a location associated with indicator 610F. In some examples, user input 634 is a deep press at a location associated with indicator 610F.

[0281] As shown in as shown in FIG. 6W, in some examples, in response to user input 634, device 600 displays multi-device interface 610 with the indicator 610F expanded to shown individual indicators for devices in the set of available playback devices as visually associated with indicator 610F. Thus, device 600 is now focused on the media session that includes the device set Living Room. As can be seen, current media interface 636 (described below) is visible (through the multi-device interface 610) in FIG. 6W. Notably, as seen in FIG. 6W, in response to input 634, device 600 ceases to display the visual indications associated with current media interface 630 and displays, instead, visual indications associated with current media interface 636. This is substitution of current media interfaces is indicative of the device changing the targeted media session from the session associated with indicator 610A to the session associated with indicator 610F (FIG. 6U).

[0282] As shown in as shown in FIG. 6X, in some examples, in response to user input 634, device 600 ceases displaying multi-device interface 610 and displays the current media interface 636. Thus, FIG. 6X illustrates an alternative response by device 600 to receiving user input 634. Current media interface 636, similar to interface 630, includes album art 636A, media identifier 636B (song title "Heart Wave"), media identifier 636C (artist "MKTS"), media identifier 636D (album "MKTS"), and device selection affordance 636E. However, each of the indicators has been updated to reflect that the media item and devices of the selected media session has changed (e.g., because device 600 is now targeted to a media session that includes a different set of devices). Notably, device selection affordance 636E includes an identifier 636F, which

identifies the currently targeted device. As shown in FIG. 6X, identifier 636F includes the text "Living Room" indicating that the current media interface 636 is a current media interface for media currently playing back on the set of devices named Living Room.

[0283] A user may desire to view the media status of the plurality of devices within their home at any time, even while viewing an interface for a non-media related application on their device. Thus, a technique for accessing a multi-device interface easily and quickly from any application executing on a device is described herein. FIGS. 6Y-6AA illustrate an exemplary technique for accessing a multi-device interface, in accordance with some embodiments. FIG. 6Y illustrates an exemplary interface for an application. Web browsing interface 640 depicts a currently displayed web browsing session (e.g., a home page of the website "Apple.com"). The web browsing interface 640 includes webpage content 640A. At FIG. 6Z, device 600 receives a user input 642. In some examples, user input 642 is a swipe gesture from an edge of a touch-sensitive surface of device 600. In some examples, in response to user input 642, device 600 displays a playback control interface. For example, in response to user input 642, device 600 displays playback control interface 606A of FIG. 6C. In some embodiments, in response to receiving user input 608A on device selection affordance 606B at playback control interface 606A, device 600 displays a multi-device interface (e.g., 610 of FIG. 6AA).

[0284] FIG. 6AA illustrates an exemplary multi-device interface visually overlaid on an application interface. As shown in FIG. 6AA, multi-device interface 610 includes individual indicators for available playback devices. In the example depicted in FIG. 6AA, there is only one media session currently active on available devices in the user's home—which includes the playback of "Same Thing" by "Jambug" on the current device 600. In some examples, the multi-device interface is concurrently displayed with previously-displayed content. For example, as shown in FIG. 6AA, the multi-device interface 610 is visually overlaid (e.g., displayed in a first region) over the web browsing interface 640 (e.g., displayed in a second region).

[0285] FIGS. 6AB-6AC illustrate exemplary techniques for adjusting a volume of available playback devices. FIG. 6AB depicts multi-device interface 610 with a single media session that includes playback of the media item "Same Thing" by "Jambug" on three devices: "iPhone" (e.g., device 600), the Kitchen set of devices, and the Living Room set of devices. FIG. 6AB

depicts individual volume control affordance 610S (corresponding to the device iPhone), individual volume control affordance 610T (corresponding to the device set Kitchen), individual volume control affordance 610U (corresponding to the device set Living Room). For example, individual volume control affordances are selectable and can be manipulated (e.g., left and right) in order to adjust volume of the respective corresponding device. In some examples, an individual volume control affordance is not displayed for devices that are not included in the current media session (e.g., the device Bedroom in FIG. 6AB does not correspond to a displayed individual volume control affordance). FIG. 6AB also depicts a group volume control affordance 610D, which is selectable and can be manipulated (e.g., left and right) in order to adjust volume of the plurality of devices included the media session represented by indicator 610A. For example, at FIG. 6AB, device 600 receives user input 644 on group volume control affordance 610D, which represents a tap and slide gesture in a leftward direction (e.g., representing a request to lower the volume). In response to the user input 644, the device lowers the group volume as represented by the location of the sliding affordance 610D, as shown in FIG. 6AC. FIG. 6AC also illustrates that the individual volume control affordances for each device in the group have been correspondingly lowered.

[0286] FIGS. 6AD-6AG illustrate exemplary interfaces for initiating playback on a preferred device. In certain instances, a user may desire to initiate playback of media automatically at one or more preferred devices. For example, because media content can include a visual aspect (e.g., if the media is a video), it can be desirable to automatically initiate playback of such content on a device with a display, or on a device with a large display (e.g., larger than the device that receives the user input request to initiate playback). For example, the user may browse for media content using their personal device (e.g., device 600) that has a relatively small screen, but desire that playback of a selected media item (e.g., a feature-length film) begin on their large-screen television. In other examples, the user may request playback of media via voice commands to a device (e.g., device 580) that does not have video playback capabilities. In this latter example, the device without the video playback capability is able to receive the user's request, but hands off video playback responsibility to another device.

[0287] Turning to the figures, FIG. 6AD illustrates an exemplary browsing interface. As shown in FIG. 6AD, media browsing interface 650 includes a media item with video content. In

this example, media item 650A represents an episode of a television show named "Super Heroes". At FIG. 6AE, the device 600 receives user input 652 associated with media item 650A. In some examples, user input 652 is a tap or touch on media item 650A. In some examples, user input 652 is a deep press on media item 650A.

[0288] In response to receiving the user input 652, device 600 displays message interface 654 in FIG. 6AF. Message interface 654 includes an indication that playback of the video content is or has been initiated on a different device ("Beginning playback on Apple TV") that is not device 600. In this example, the device named "Apple TV" is different than device 600, and represents a preferred device for playback of video media. For example, the device Apple TV can be connected to the user's large-screen TV. In some embodiments, in response to the user input (e.g., user input 652) the device (e.g., 600) determines whether it is currently connected to a home network. In accordance with a determination that the device is currently connected to the home network, the device initiates (e.g., transmits an instruction to initiate) playback on the preferred device. For example, if the device determines that the user is home, it forgoes initiating playback on itself, and initiates playback on the external preferred device. In accordance with a determination that the device is not currently connected to the home network, the device initiates playback on itself. For example, if the device determines that the user is not home, it initiates playback on itself. In some examples, the preferred device is a set of preferred devices (e.g., one or more). In some examples, a preferred device (or set of devices) is predetermined. For example, a preferred device can be selected by the user and saved into device or user account settings. In another example, a preferred device is automatically selected. For example, if the user's home only includes one recognized video capable external device (e.g., an Apple TV), it may be designated as the preferred device. In some embodiments, the device prompts the user whether to initiate playback on a preferred device. For example, regardless of whether a preferred device has been preconfigured by the user, the device 600 may prompt (e.g., visually or audibly) and ask "Would you like to initiate playback of "Super Heroes" on your Apple TV?" in response to user input 652.

[0289] FIG. 6AG illustrates an exemplary playback control interface for video content. Playback control interface 656 includes an indication 656A of the currently playing media content ("Super Heroes"), playback control affordances 656B (e.g., rewind 10 seconds, pause,

and skip 10 seconds), and an indication 656C of the device currently playing back the media content ("Apple TV"). In some embodiments, interface 656 is displayed subsequent to message interface 654.

[0290] FIGS. 6AH-6AQ illustrate exemplary multi-device interfaces. As described herein, a multi-device interface can be used to view and control media playback status of a plurality of available playback devices throughout a set of interconnected devices (e.g., devices in a user's home and associated with the user's account). Thus, a plurality of different types of devices can be controlled and each device can have a different interface. For example, the set of devices in a home can comprise a first smart speaker (e.g., 580) made by a first manufacturer, a second smart speaker (e.g., 580) made by a second manufacturer different than the first, and home media hub (e.g., an Apple TV by Apple Inc. of Cupertino, California USA). An exemplary home media hub can include or otherwise be connected to a large display (e.g., larger than the display of personal device 600), and thus can have an identical, a similar, or a different interface than those described above with respect to device 600. Additional exemplary interfaces are described in detail below.

[0291] FIG. 6AH illustrates an exemplary current media interface 664, displayed on display 660A of device 660. In some embodiments, device 660 includes one or more features of device 100, 300, or 500. Device 660 is itself a device included in the set of devices associated with the user's home (or is otherwise configured to act as a display for a device included in the set of devices associated with the user's home). Current media interface 664 includes a device selection affordance 664A (e.g., that functions similar or the same as device selection affordance 630E of FIG. 6S). Device selection affordance 664A is associated with the text "Living Room Apple TV", which indicates that the device 660 is currently targeted at itself, device 660. Current media interface 664 also includes an indication of currently playing media (e.g., indicator 664C), which is the song titled "Heart Wave" (e.g., indicator 664D) by artist "MKTS" (e.g., indicator 664E). In some examples, indicator 664C of currently playing interface 664 is an animated indictor, further communicating that the associated media item is currently playing back. Indicators 664F and 664G correspond to one or more other media items available for playback.

[0292] FIG. 6AH also illustrates an exemplary remote control for controlling device 660. In some examples, remote control 662 is connected to device 660 (e.g., paired) and used by a user to interact with interfaces displayed on device 660. Because a remote control is effectively an extension of a device that is used for controlling the device remotely, for the purposes of this disclosure, a user input received at device 662 is considered the same as a user input received at device 660, unless otherwise noted herein. For example, a tap input received on device 662 that represents a request to select an affordance is considered the same as a tap on device 660 (e.g., if device 660 included a touch-sensitive surface) representing a request to select the same affordance, unless otherwise noted. However, one of skill in the art would appreciate that a user input can be received at any device and communicated to another device (e.g., for processing and performing an action in response), and that the particular device that physically receives the user input is not important, unless specifically noted. Likewise, in the example that device 660 is an external display device (e.g., a television) used by another device (e.g., an Apple TV that is a "home device" that receives and processes user input playback commands), because device 660 is effectively an extension of the other device, device 660 as referred to herein interchangeably with said other device, unless otherwise noted. For example, a reference to the "focus" or "target" of device 660 with respect to controlling media refers to the focus or target of the other device (e.g., Apple TV), unless otherwise noted.

[0293] In FIG. 6AH, device 662 receives user input 668 on device 662. User input 668 represents selection of the device selection affordance 664A. In some examples, user input 668 is a button press on device 662 while device selection affordance 664A is highlighted or otherwise selected at interface 664. In some examples, user input 662 is a tap, a tap and hold, or a deep press on a touch-sensitive surface. In some embodiments, device 660 displays a multi-device interface (e.g., 670 of FIG. 6AI) in response to receiving user input 668.

[0294] FIG. 6AI illustrates an exemplary multi-device interface, in accordance with some embodiments. Multi-device interface 670 includes an identification of the current focus (e.g., the currently-selected media session) of device 660. Header 670A indicates that the current focus of the device is the Living Room Apple TV (e.g., device 660). Multi-device interface 670 also includes a first indicator 670B, which indicates the currently-targeted media session. For example, the media session represented by indicator 670B includes playback of the song titled

"Heart Wave" by artist "MKTS" from album "MKTS". Together, header 670A and indicator 670B (which may be collectively referred to as an indicator) indicate the currently-targeted media session (e.g., media playing back, and on which devices) of interface 670. Multi-device interface 670 also includes a second indicator 670C, which corresponds to a different media session than the currently-selected media session. As shown in FIG. 6AI, indicator 670C is displayed (e.g., partially displayed) on the edge of the display.

[0295] As described above with respect to multi-device interface 610, multi-device interface 670 includes individual indicators representing available playback devices (e.g., devices in the user's home). Indicator 670D identifies the current device (named "Living Room Apple TV") and includes the selection indicator 670K. In this example, selection indicator 670K includes a checkmark, which indicates that the device 670D is currently playing back the media of indicator 670B (e.g., is included in the currently-targeted media session). Indicator 670E identifies the set of devices named "Kitchen" and includes indicator 670L, which does not include a checkmark (e.g., the Kitchen set of devices is not currently playing back the media track "Heart Wave" by "MKTS"). Indicator 670E also includes an indication that the Kitchen set of devices are currently playing back other media. For example, in FIG. 6AI, the indicator 670E includes the title of the track "Same Thing" and the artist "Jambug". Indicator 670F identifies the set of devices named "Bedroom" and includes indicator 670M, which does not include a checkmark (e.g., the Bedroom set of devices is not currently playing back the media track "Heart Wave" by "MKTS"). Indicator 670F also includes an indication that the Bedroom set of devices are currently playing back other media. For example, in FIG. 6AI, the indicator 670F includes the title of the track "Vibes" and the artist "John Appleseed". Indicator 670G identifies the set of devices named "Patio" and includes indicator 670N, which does not include a checkmark (e.g., the Patio set of devices is not currently playing back the media track "Heart Wave" by "MKTS").

[0296] While multi-device interface 670 is displayed, device 662 receives user input 672, as shown in FIG. 6AI. In some examples, user input 672 is a swipe gesture on a touch-sensitive surface of device 662. For example, user input 672 is a leftward swipe.

[0297] In response to a user input (e.g., swipe gesture), in some examples, the device updates multi-device interface 670A to focus on a different media session. As shown in FIG. 6AJ, in response to user input 672 that is a leftward swipe, device 660 replaces display of the first indicator 670B with the display of second indicator 670C, which had been displayed immediately to the right of the first indicator 670B in FIG. 6AI. Thus, the multi-device interface 670 can be used to switch between different media sessions, visually indicated by switching between corresponding indicators. In some examples, other user inputs can be used to switch the multi-device interface between media sessions. For example, other user inputs include up or down directional gestures, left or right directional gestures, selection of one or more affordances (e.g., a left or right arrow), press of a button on a device (e.g., device 662), or any other appropriate type of user input.

[0298] As shown in FIG. 6AJ, in response to the user input 672, the current media session now corresponds to indicator 670C, which includes playback on the set of devices identified as "Kitchen" (e.g., as shown in updated header 670A) of the media item titled "Same Thing" by artist "Jambug" from the album "Fly Away" (e.g., as shown in indicator 670C). Additionally, the individual device indicators 670D-670G have been updated based on the currently displayed media session associated with indicator 670C. Thus, the device set named Kitchen corresponding to indicator 670E is the only device set that includes a checkmark (e.g., in corresponding indicator 670L). Further, FIG. 6AJ shows indicator 670H (not previously displayed) is now displayed in the area of the display where indicator 670C was displayed prior to user input 672. Thus, an additional leftward swipe (e.g., 674 of FIG. 6AJ), in this example, would cause indicator 670H to replace indicator 670C. FIG. 6AJ also shows indicator 670B displayed (e.g., partially) to the left of indicator 670C. Thus, a rightward swipe, in this example, would cause indicator 670B to replace display of indicator 670C (e.g., as shown in FIG. 6AI).

[0299] FIG. 6AK illustrates multi-device interface 670 with a current media session corresponding to indicator 670I (e.g., one leftward swipe from the media session represented by indicator 670H, and two leftward swipes from the media session represented by indicator 670C). In some embodiments, a media session corresponds to a set of devices that are not currently playing back media. For example, as shown in FIG. 6AK, the current media session includes the device set named "Patio". Rather than include an identification of a currently playing media

item, indicator 670I includes the text "Choose Music". Likewise, indicator 670G does not include an indication of a media item.

[0300] FIG. 6AL illustrates multi-device interface 670 with a current media session corresponding to indicator 670H (e.g., one leftward swipe from the media session represented by indicator 670C, and one rightward swipe from the media session represented by indicator 670I). As shown in FIG. 6AL, indicator 670H is currently highlighted. In this example, while the indicator 670H is highlighted, device 662 receives a user input 676. In some examples, user input 676 is a tap, touch, or press received at device 662. In response to receiving the user input 676, device 660 displays current media interface 664 as shown in FIG. 6AM. As shown in FIG. 6AM, current media interface 664, displayed in response to user input 676, indicates the device 660 is currently targeted at a media session that includes the devices named "Bedroom" (e.g., as indicated by device selection affordance 664A) that is currently playing the song titled "Vibes" by "John Appleseed" (e.g., corresponding to indicator 664F, showing the album cover for the song). Thus, the focus of the device 660 has been changed (e.g., as compared to FIG. 6AH), and selection of new media for playback (e.g., using interface 664) will affect the currently targeted media session.

[0301] FIG. 6AN illustrates an exemplary multi-device interface that includes a currently-selected media session that includes two device sets. In FIG. 6AN, both the Kitchen and Living Room sets of devices are included in the current media session (e.g., playback of the song titled "Same Thing"). As shown in FIG. 6AN, indicator 670E is currently highlighted. User input 678 is received at device 662 while indicator 670E is highlighted.

[0302] In some examples, in response to user input 678 on a selected indicator corresponding to one or more devices, a volume control affordance is displayed. As shown in FIG. 6AO, in response to the user input 678 while indicator 670E is selected, device 660 enters a volume adjustment mode for the corresponding device set named Kitchen. For example, a volume adjustment affordance 670J is displayed (e.g., replacing indicator 670E temporarily). While in the volume adjustment mode, the device 662 receives a user input 679, and in response, adjusts the volume of the Kitchen device set accordingly. In some examples, user input 679 is a directional swipe (e.g., left or right, up or down).

[0303] FIG. 6AP illustrates an exemplary browser interface, in accordance with some embodiments. As shown in FIG. 6AP, browser interface 680 includes an indicator 680A of media that is now playing, which represents the currently-targeted media session, as well as recently added media items (e.g., that when selected cause playback to initiate on the devices in the currently-selected media session). In some examples, while browser interface 680 is displayed, user input 681 is received. In some examples user input 681 is a selection of button 682 of device 662. In some examples, user input 681 is a tap and hold user input, a press and hold (e.g., of a button), a deep press, a double tap or double press, or any other appropriate user input on device 662 or device 660.

[0304] As shown in FIG. 6AQ, in response to user input 681, device 660 displays multi-device interface 670 overlaid on the browser interface 680. In some examples, multi-device interface 670 is displayed concurrently with other interfaces of device 660. For example, user input 681 provides a quick access to multi-device interface so that a user can quickly view and control the media playback status of devices throughout their home, even when performing other tasks or operations (e.g., using other, non-media related applications).

[0305] FIG. 6AR illustrates a variety of exemplary playback control interfaces. Playback control interface 690A represents an interface for controlling audio playback, in accordance with some embodiments. Playback control interface 690B represents an interface for controlling video playback, in accordance with some embodiments. Playback control interface 690C represents an interface for controlling audio, in accordance with some embodiments. For example, playback control interface 690C is suited for use as a pop-up interface (e.g., of a tool bar or web browser). For instance, in response to selection of an affordance, interface 690C can protrude from the selected affordance (e.g., from the pointed shape on one of its sides), temporarily covering currently displayed content (e.g., until dismissed). Playback control interface 690D is an alternative to 690A, for example, and can be used as a persistent "widget" in a device operating system interface.

[0306] In some embodiments, a device (e.g., device 600) is configured to receive playback coordination responsibility from another device (e.g., device 580). For example, a device that is configured to receive or perform playback coordination responsibility (also referred to as a "hero

device") is a device that is configured to perform one or more of the following tasks: issue commands to one or more external devices to synchronize and/or control playback of media on the external devices; store a media item or a playback queue for playback; access a content stream from a remote location (e.g., an internet location); provide a content stream (e.g., of media) to one or more external devices for playback; and process playback control instructions for a media session (e.g., receive, process (e.g., de-duplicate redundant requests), and forward instructions to appropriate devices). In some embodiments, device 600 coordinates playback on a set of device. In some embodiments, the set of devices excludes the device 600. For example, if the set of devices does not include a device configured to be hero device, device 600 retains playback coordination responsibility. In some embodiments, device 600 transfers playback coordination responsibility to a device configured to be a hero device. For example, if the set of devices includes a device configured to be hero device, device 600 transfers playback coordination responsibility to that device (e.g., such as a device 580). Playback coordination responsibility is described in more detail below.

[0307] FIG. 7 is a flow diagram illustrating a method for accessing media playback status of a plurality of available playback devices using an electronic device in accordance with some embodiments. Method 700 is performed at a device (e.g., 100, 300, 500, 580) with a display. Some operations in method 700 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0308] As described below, method 700 provides an intuitive way for accessing media playback status of a plurality of available playback devices. The method reduces the cognitive burden on a user when accessing media playback status of a plurality of available playback devices, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to access media playback status of a plurality of available playback devices faster and more efficiently conserves power and increases the time between battery charges.

[0309] The electronic device (e.g., 600) receives (702) a user input. For example, the device receives user input (e.g., 608A of FIG. 6E) selection of an affordance (e.g., 606B of FIG. 6E). In

some examples, the device receives a user input (e.g., 608B of FIG. 6F) associated with a playback control interface (e.g., 606A of FIG. 6C).

[0310] In response to receiving the user input, the electronic device (e.g., 600) displays (704), on the display, a multi-device interface (e.g., 610) that includes: one or more indicators (e.g., 610A, 610F, 610G) associated with a plurality of available playback devices that are connected to the device (e.g., 600) and available to initiate playback of media from the device (e.g., 600), and a media playback status (e.g., as indicated by 610F and 610G) of the plurality of available playback devices. In some examples, a media playback status includes one or more of: an indication of what is currently being played on the plurality of available playback devices. In some examples, the devices in the plurality of available playback devices are associated with a user, a user account, or a plurality of devices associated with a user. For example, the devices may be a collection of devices belonging to the user or designated as residing inside of a user's home. In some embodiments, an indicator of the one or more indicators can be associated with more than one available playback device. For example, devices that are currently playing the same content can be grouped and displayed as a single indicator (e.g., 610A of FIG. AB).

[0311] Displaying a multi-device interface that includes one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, along with a media playback status of the plurality of available playback devices, provides the user with visual feedback about the state of a plurality of devices that are available for media playback. Providing improved visual feedback to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0312] Additionally, displaying such a multi-device interface allows the user to collectively view a plurality of states for a plurality of available devices in one optimized interface. Providing an optimized interface to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces

power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0313] In some embodiments, one or more devices are "connected" when data can be exchanged between the devices. For example, devices can be connected to a common Local Area Network ("LAN") that enables data exchange between the devices. In some embodiments, one or more connected devices are connected directly. In some examples, a direct connection between two devices does not rely on an intermediate device between two directly connected devices. For example, a direct connection can be formed between a device that creates a Wi-Fi connection directly to a second device, without requiring an intermediate wireless access point ("WAP") to facilitate the connection.

[0314] In some embodiments, one or more devices are connected indirectly. In some examples, an indirect connection between two devices includes (e.g., is routed through) an intermediate device between two connected devices. For example, an indirect connection can be formed between a device that is connected to a common wireless (e.g., Wi-Fi) connection through an intermediate WAP.

[0315] In some embodiments, one or more devices are "connected" by virtue of each being logged into (or otherwise associated with) the same user account. For example, a user may have two personal devices that are both logged into an account for a cloud-based service and thus are connected. For instance, the user's phone (e.g., an iPhone) and their tablet computer (e.g., an iPad) can both be logged into the user's iCloud account. In this example, the iPhone and the iPad can exchange data through the cloud-based service (e.g., a cloud server). In some embodiments, the device is connected to the available playback devices such that the device can initiate playback of media on one or more of the available playback devices. For example, an available playback device can begin playback of an media file (or stream), transfer playback (currently playing on the device) to the available device, or add the available device to concurrently play back of media a currently playing on the device).

[0316] In some embodiments, the user input (e.g., 608A of FIG. 6E) is a first user input, and the electronic device (e.g., 600), receives a second user input (e.g., 615 or 616) representing selection of the one or more indicators associated with the plurality of available playback

devices. In response to receiving the second user input, the electronic device transmits an instruction to initiate playback on the plurality of available playback devices. In some embodiments, the plurality of devices begin playback immediately. In some embodiments, the available devices begin playback without building up a buffer of media data and then, during playback, build up a buffer. Thus, for example, delay between when a user selects a media item and hears playback is reduced. In some embodiments, playback does not begin until a certain proportion of devices of the plurality are ready to begin playback. For example, playback can begin when 75% of devices are ready (e.g., have buffered a minimum amount of media).

[0317] In some embodiments, the user input (e.g., 608A of FIG. 6E) is a first user input, and while displaying the multi-device interface (e.g., 610), the electronic device (e.g., 600) displays, on the display (e.g., 602), a first indicator (e.g., 610E, 610F, or 610G of FIG. 6H). The electronic device receives a third user input (e.g., 612 or 616) representing selection of the first indicator (e.g., 610E or 610F). In response to receiving the third user input, the electronic device displays, on the display, affordances (e.g., 610H-610O of FIG. 6J or FIG. 6N), for each of the devices of the plurality of available playback devices. For example, after selecting an indicator (e.g., 610) for a group of devices (e.g., "Kitchen" and "Living Room" of FIG. 6N), the multi-device interface includes affordances (e.g., 610H, 610J, 610L, and 610N) for each device in the home (e.g., iPhone, Bedroom, Kitchen, and Living Room in the example in FIG. 6N).

[0318] In some embodiments, while the electronic device (e.g., 600) is playing back a first media item (e.g., represented by 610A of FIG. 6J), the electronic device receives a fourth user input (e.g., 616) representing selection of a second indicator (e.g., 610F) of the one or more indicators, wherein the second indicator is associated with a set of playback devices that does not include the device. In some embodiments, a set of playback devices includes one or more playback devices. Subsequent to receiving the fourth user input representing selection of the second indicator, the electronic device displays, on the display, a browsing interface (e.g., 620), wherein the browsing interface includes an affordance (e.g., 620B) associated with a second media item. The electronic device (e.g., 600) receives a fifth user input (e.g., 622) representing selection of the affordance associated with the second media item. In response to receiving the fifth user input, the electronic device (e.g., 600): transmits an instruction to initiate playback of the second media item on the set of playback devices that does not include the device (e.g., 600);

and continues playing back the first media item (e.g., represented by 610A of FIG. 6J) on the electronic device (e.g., 600). For example, in response to the user input selection of affordance 620B (associated with media item "Monday Morning" by "The Fomoers"), the device initiates playback of the associated media item on the selected devices named "Kitchen" and "Living Room" (e.g., a shown selected in FIG. 6N). Further, playback continues on the electronic device (e.g., named iPhone in this example) of the media item titled "Same Thing" by artist "Jambug" (e.g., as shown in FIG. 6N).

[0319] In some embodiments, the browsing interface (e.g., 620) includes an affordance associated with a third media item (e.g., 620B), and the electronic device receives a sixth user input (e.g., 622) associated with the affordance associated with the third media item. In accordance with receiving the sixth user input, the electronic device displays, on the display, the multi-device interface. For example, after receiving a deep press on a media item (e.g., a song), the device displays the multi-device interface (e.g., 610 of FIG. 6J).

[0320] In some embodiments, the sixth user input (e.g., 622) is a press and hold. In some embodiments, in response to the sixth user input, the electronic device displays a menu (e.g., 624) with an affordance (e.g., 624A), selection of which causes the multi-device interface (e.g., 610) to be displayed.

[0321] In some embodiments, the sixth user input (e.g., 622) has a characteristic intensity. In accordance with a determination that the characteristic intensity of the sixth user input meets the intensity-dependent criteria, and subsequent to receiving the sixth user input, the electronic device (e.g., 600) displays, on the display (e.g., 602), the multi-device interface (e.g., 610). In some embodiments, in response to the sixth user input, the electronic device displays a menu (e.g., 624) with an affordance (e.g., 624A), selection of which causes the multi-device interface (e.g., 610) to be displayed. In accordance with a determination that the characteristic intensity of the sixth user input does not meet the intensity-dependent criteria, and subsequent to receiving the sixth user input: the electronic device forgoes displaying, on the display, the multi-device interface; and transmits an instruction to initiate playback of the third media item (e.g., 620B) on the set of playback devices that does not include the device (e.g., Living Room + Kitchen, as shown in FIG. 6N).

[0322] In some embodiments, to display the multi-device interface (e.g., 610), the electronic device (e.g., 600) concurrently displays, on the display (e.g., 602): a first region (e.g., 611A of FIG. 6U) that includes the multi-device interface (e.g., 610 of FIG. 6U); and a second region (e.g., 611B of FIG. 6U) that includes content (e.g., 630) displayed prior to displaying the multi-device interface.

[0323] In some embodiments, the first region (e.g., 6611A) is visually overlaid on the second region (e.g., 611B, as shown in FIG. 6U). In some embodiments, the second region is visually blurred, (e.g., to indicate that the first region is an active window). In some embodiments, the second region appears darkened (e.g., to indicate that the first region is an active window).

[0324] In some embodiments, the electronic device (e.g., 600) receives selection (e.g., user input 634) of an indicator (e.g., 6610F of FIG. 6V) of the one or more indicators, wherein the indicator includes a representation of a fourth media item (e.g., "Heart Wave" as shown in FIG. 6V). In some examples, the representation of the fourth media item is a thumbnail image that includes album art, an artist thumbnail, or the like. In response to receiving selection of the indicator, and while continuing to concurrently display the first and second regions, the electronic device (e.g., 600) updates display of the second region (e.g., 611B) to include at least a portion of the representation of the fourth media item (e.g., as shown in FIG. 6W). For example, the second region can be updated to reflect that the device has received user selection of a different media item than that which is currently playing back, a representation of which was displayed immediately prior to receiving the selection. Thus, the electronic device visually indicates that it has been retargeted to the devices that are playing back the fourth media item. In some examples, the second region represents a media application on the device. For example, a media application includes any application executing on the device usable for browsing, selecting, or otherwise initiating playback of media items.

[0325] In some embodiments, the user input (e.g., 608A or 608B) is a first user input, and prior to receiving the user input, the electronic device (e.g., 600) receives a seventh user input (e.g., 604 or 642). In response to receiving the seventh user input, the electronic device displays, on the display (e.g., 602), a playback control interface (e.g., 606A of FIG. 6C) that includes: one or more playback control affordances (e.g., 606M of FIG. 6C); a playback device selection

affordance (e.g., 606B of FIG. 6C); and an affordance for controlling a data connection of the device (e.g., 606N). The electronic device receives the first user input (e.g., 608A of FIG. 6E, or 608B of FIG. 6F), wherein the first user input is associated with the playback control interface (e.g., selection of 606B of FIG. 6E, or a deep press of 606A of FIG. 6F).

[0326] In some embodiments, the electronic device receives the first user input (e.g., 608A or 608B), and the first user input has an input characteristic. In some embodiments, an input characteristic is a characteristic intensity of an input. In some embodiments, an input characteristic is a length of time associated with an input. In some examples, a length of time associated with an input is a length of time that a touch input was continuously detected on a touch-sensitive surface (e.g., of device 600).

[0327] The electronic device (e.g., 600) determines whether the input characteristic meets an input characteristic-dependent criteria. In some embodiments, to determine whether the input characteristic meets an input characteristic-dependent criteria, the electronic determines whether the characteristic intensity meets an intensity-dependent criteria. For example, whether the characteristic input of the input is above a threshold intensity. In some embodiments, to determine whether the input characteristic meets an input characteristic-dependent criteria, electronic device determines whether the user input meets a time-based criteria. For example, whether the touch is detected (e.g., on a touch-sensitive surface associated with the electronic device) for a predetermined amount of time. Input characteristics and input characteristic dependent criteria other than those based on intensity or time can be used. In accordance with a determination that the input characteristic meets the input characteristic-dependent criteria, the device displays, on the display, the multi-device interface. In accordance with a determination that the input characteristic intensity does not meet the input characteristic-dependent criteria, the electronic device forgoes displaying, on the display, the multi-device interface. In some embodiments, further in accordance with a determination that the input characteristic intensity does not meet the input characteristic-dependent criteria, the electronic device determines a location associated with the first user input. In accordance with a determination that the location is associated with a playback control affordance (e.g., 606M of FIG. 6C), the device performs a corresponding playback control function. For example, if the first user input does not have a

characteristic intensity that exceeds a threshold intensity, and is received at a location associated with a pause playback affordance, then the device pauses playback.

[0328] In some embodiments, the seventh user input (e.g., 642) is received while displaying a user interface of an application executing on the device (e.g., 640 of FIG. 6Y). While continuing to display at least a portion (e.g., 640 of FIG. 6AA) of the user interface of the application, the electronic device (e.g., 600) displays, on the display (e.g., 602), the playback control interface (e.g., 606A). Subsequent to receiving the first user input, and while continuing to display the at least a portion of the user interface of the application, the electronic device (e.g., 600) displays, on the display, the multi-device interface (e.g., 610 of FIG. 6AA). In some embodiments, a user input (e.g., a touch or click) outside of the multi-device interface portion (e.g., outside of 610 of FIG. 6AA) causes the electronic device to cease displaying the multi-device interface (e.g., and display 640 as shown in FIG. 6Y). For example, the multi-device interface ceases being displayed overlaid on the application and the application (e.g., represented by 640 of FIG. 6Y) becomes the active user interface again.

[0329] In some embodiments, the multi-device interface (e.g., 610 of FIG. 6AB) includes a volume control affordance (e.g., 610D of FIG. 6AB) associated with the plurality of available playback devices (e.g., the devices represented by 610H, 610L, and 610N). The electronic device receives an eighth user input (e.g., 6644) associated with the volume control affordance. In response to receiving the eighth user input, the device transmits an instruction to adjust a playback volume of the plurality of available playback devices (e.g., the devices represented by 610H, 610L, and 610N). For example, the device (e.g., 600) transmits instructions to the devices "Kitchen" and "Living Room" that are currently targeted by the device and are playing back the same media content.

[0330] In some embodiments, the multi-device interface includes a plurality of individual volume control affordances (e.g., 610S, 610T, and 610U of FIG. 6AB), each associated with a single device of the plurality of available playback devices (e.g., the devices represented by 610H, 610L, and 610N, respectively). The device receives a ninth user input associated with an individual volume control affordance (e.g., 610U). For example, the electronic device receives a selection and movement of the slider 610U in FIG. AB. In response to receiving the ninth user

input, the device transmits an instruction to adjust a playback volume of the associated device (e.g., device represented by 610N), wherein the associated device is different than the electronic device (e.g., 600). For example, the electronic device transmits an instruction to adjust the volume for the device associated with the affordance 610U, forgoing adjustment of volume for other devices in the group.

[0331] In some embodiments, the one or more indicators associated with the plurality of available playback devices includes: a first indicator (e.g., 670B of FIG. 6AI) representing a fifth media item currently playing on the electronic device (e.g., 660), wherein the first indicator is displayed in a first region (e.g., 670A of FIG. 6AI); a second indicator (e.g., 670C of FIG. 6AI) representing a group of devices (e.g., 670E of FIG. 6AJ) that are each currently playing back a sixth media item; and individual indicators (e.g., indicators 670D, 670E, 670F, and 670G of FIG. 6AI) for each device of the plurality of available playback devices. While displaying the multi-device interface, the electronic device (e.g., 600) receives a tenth user input (e.g., 672) representing a request to replace display of the first indicator (e.g., 670B of FIG. 6AI) in the first region (e.g., 670A of FIG. 6AI) with display of the second indicator (e.g., 670C). In response to the tenth user input, the electronic device replaces display (e.g., as illustrated in FIG. 6AJ) of the first indicator (e.g., 670B) in the first region (e.g., 670A) with display of the second indicator (e.g., 670C as illustrated in FIG. 6AJ). In some embodiments, the tenth user input is received at a second device (e.g., 662) associated with the electronic device (e.g., 600 or 660).

[0332] In some embodiments, the electronic device (e.g., 660) displays an indication (e.g., 670K) of whether each device of the plurality of available playback devices is currently playing back a media item associated with an indicator displayed in the first region.

[0333] In some embodiments, the electronic device (e.g., 660) receives an eleventh user input (e.g., 678) associated with an individual indicator (e.g., 670E of FIG. 6AN) for a device of the plurality of available playback devices. The electronic device determines whether the eleventh user input represents a selection that exceeds a threshold duration. For example, the input can be a tap and hold on a touch-sensitive surface of the device or a touch-sensitive surface associated with the device (e.g., of 662). In accordance with the eleventh user input representing a selection that exceeds the threshold duration, the electronic device enters a volume adjustment

mode for the device associated with the individual indicator (e.g., illustrated in FIG. 6AO). In some examples, entering the volume adjustment mode can include displaying a visual indication (e.g., 670J of FIG. 6AO) that the electronic device is in a volume adjustment mode. For the visual indication can indicate which device or devices are currently subject to volume adjustment in the volume adjustment mode. For example, indicators for other devices can be grayed out or otherwise altered in appearance in response to the eleventh user input. In some examples, a volume indicator (e.g., 670J) such as a slider or knob indicating a volume level appears, appears enlarged, or is otherwise visually altered so as to visually communicate that the electronic device is currently in the volume adjustment mode. In accordance with the eleventh user input (e.g., 678) not representing a selection that exceeds the threshold duration, forgoing entering the volume adjustment mode for the device associated with the individual indicator.

[0334] In some embodiments, the electronic device (e.g., 600, 660) receives a twelfth user input representing a request to initiate media playback on the plurality of available devices. In response to receiving the twelfth user input: in accordance with a determination that the plurality of available devices includes at least one device configured to receive playback coordination responsibility, the electronic device transfers playback coordination responsibility to the at least one device of the plurality of devices. In accordance with a determination that the plurality of available devices does not include at least one device configured to receive playback coordination responsibility, the electronic device (e.g., 600, 660) coordinates the media playback on the plurality of devices.

[0335] In some embodiments, while the electronic device (e.g., 600 or 660) is coordinating the media playback on the plurality of devices, the electronic device receives a thirteenth user input representing a request to initiate a phone call on the device. For example, the device receives user input representing a request to answer a phone call or to place a phone call. In response to receiving the thirteenth user input representing the request to initiate the phone call, the electronic device: initiates the phone call; and continues coordinating the media playback on the plurality of devices during the phone call.

[0336] In some embodiments, further in response to receiving the twelfth user input, the electronic device (e.g., 600, 660) transmits data representing a playback queue to each of the devices in the plurality of devices.

[0337] In some embodiments, the electronic device (e.g., 600 or 660) receives a fourteenth user input representing a request to initiate media playback on the plurality of available devices, wherein the fourteenth user input is associated with a first user. In some examples, the electronic device determines that a user input is associated with the first user based on detected voice input, based on a device that sent the request, or the like. In response to receiving the fourteenth user input, in accordance with a determination that the first user is subject to a playback restriction, the electronic device (e.g., 600, 660) prevents the electronic device (e.g., 600, 660) from initiating media playback on the plurality of available devices in response to the eleventh user input. In accordance with a determination that the first user is not subject to the playback restriction, the electronic device allows the device to initiate media playback on the plurality of available devices in response to the eleventh user input.

[0338] In some embodiments, the electronic device (e.g., 600 or 660) receives a fifteenth user input (e.g., 652 of FIG. 6AE) representing a request for media output on the device. In some embodiments, the fifteenth user input represents a request to initiate media playback on the device. In some embodiments, the fifteenth user input represents a request to output media having a visual component. For example, media having a visual component includes, but is not limited to, images, videos, browser applications, electronic messages (e.g., emails, text messages), or the like. In response to receiving the fifteenth user input: in accordance with a determination that the request for media output is a request to display content (e.g., represented by 650A of FIG. 6AE), the electronic device transmits an instruction to a preferred device (e.g., 660 of FIG. 6AH) of the plurality of available playback devices to initiate display of the content in response to the twelfth request. In some embodiments, request for media output is a request to playback content having a visual component (e.g., video). For example, video content includes but is not limited to: a movie, a show, a video stream, or the like. In some embodiments, a preferred device is a device that is capable of outputting video (e.g., an Apple TV), or a device connected to a large display (e.g., predefined as preferred). In accordance with a determination that the request for media output is not a request to display content, the electronic device (e.g.,

600 or 660): forgoes transmitting the instruction to the preferred device of the plurality of available playback devices to initiate display of the content in response to the twelfth request; and initiates media output on the device in response to the twelfth user input. For example, if the requested playback is of audio content, the electronic device plays back the audio on the requested device instead.

[0339] Note that details of the processes described above with respect to method 700 (e.g., FIG. 7) are also applicable in an analogous manner to the methods described below. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 700. For brevity, these details are not repeated below.

[0340] FIGS. 8A-8P illustrate exemplary user interfaces for accessing a multi-device interface, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 9.

[0341] FIGS. 8A-8B illustrate an exemplary current media interface, in accordance with some embodiments. At FIG. 8A, device 800 displays, on display 802, current media interface 804. Current media interface 804 (also referred to as a "now playing interface") identifies a media item that is currently being played back. In some embodiments, device 800 includes one or more features of device 100, 300, or 500. Current media interface 804 includes album art 804A, media identifier 804B (song title "Same Thing"), media identifier 804C (artist "Jambug"), media identifier 804D (album "Fly Away"), and device selection affordance 804E. As shown in FIG. 8A, the device selection affordance 804E does not include an identifier for another device. Thus, in this example, the media item identified in the current media interface 804 is currently playing back on the current device, device 800. In some examples, the media identified by the currently media interface is currently being played back on other devices as well (e.g., in addition to the current device, or a set that excludes the current device). As shown in FIG. 8B, device 800 receives user input 806 representing selection of device selection affordance 804E. In some examples, user input 806 can be a tap or touch on a location associated with the displayed device selection affordance 804E.

[0342] FIG. 8C illustrates an exemplary multi-device interface, in accordance with some embodiments. In some embodiments, device 800 displays multi-device interface 810 in response to user input 806. As shown in FIG. 8C, multi-device interface 810 includes indicators 810A-810K. Multi-device interface 810 is analogous to multi-device interface 610, described above (e.g., with respect to FIG. 6J). Likewise, indicators 810A-810K are analogous to the indicators of multi-device interface 610. Accordingly, the above description of a multi-device interface (e.g., 610) is hereby incorporated by reference with respect to multi-device interface 810.

[0343] In particular, FIG. 8C depicts a multi-device interface 810 that includes an indicator 810J that represents a media session that includes the device set "Kitchen + Living Room", which is comprised of the device set "Kitchen" and the device set "Living Room". In this example, the Kitchen set represents a single smart speaker device (e.g., device 580), and the Living Room set represents a different single smart speaker device (e.g., another instance of device 580). The media session represented by indicator 810J includes playback of the media item titled "Heart Wave" by artist "MKTS" on the device set "Kitchen + Living Room".

[0344] FIGS. 8D-8H illustrate exemplary techniques for refocusing a device, in accordance with some embodiments. At FIG. 8D, the device 800 receives user input 812 representing selection of indicator 810J. In some embodiments, user input 812 is a tap or touch on a location associated with the indicator 810J. In some embodiments, user input 812 is a deep press input or a tap and hold input on a location associated with the indicator 810J. As shown in FIG. 8D, the current media interface 804 is visible on display 802 concurrently with the display of multi-device interface 810. Specifically, device 800 displays multi-device interface 810 in a first region 828A that is visually overlaid on a second region 828B. Device 800 displays the current media interface 804 in the second region 828B. As shown in FIG. 8D, second region 828B (e.g., the current media interface 804) is partially obscured by the first region 828A, but it still visible. Thus, device 800 provides a visual indication (e.g., other than the multi-device interface 810) of the current focus of the device. For instance, because the album art displayed with the indicator 810A matches the album art of current media interface 804, the device indicates that the current focus of the device, or the focus of a particular application (e.g., associated with the current media interface 804), is the current device "iPhone", which is currently playing the media item associated with the matching album art.

[0345] In response to receiving user input 812, device 800 displays an expanded version of indicator 810J, as shown in FIG. 8E. The expanded indicator 810J (and the reduced size of indicator 810A) of FIG. 8E provide visual indication that the focus of the device is now on the media session represented by indicator 810J. As shown in FIG. 8E, in response to changing the focus: indicator 810B no longer includes a checkmark in indicator 810F (e.g., because the iPhone is not part of the currently-selected media session), indicators 810D and 810E now include checkmarks in indicators 810H and 810I, respectively (e.g., because the Kitchen and Living Room set of devices are part of the currently-selected media session). In some examples, an indicator for one or more devices indicates that the corresponding one or more devices are unavailable. For example, an unavailable device is a device that is: offline, currently installing a software or firmware update, configured to be unavailable (e.g., via a setting or button), or the like. In some examples, a corresponding indicator is not displayed for unavailable devices. Alternatively, an indicator for such device or devices is displayed, but with an indication that the device is unavailable for media playback. For example, FIG. 8E depicts indicator 810C, which includes the text "Updating" indicating that the Bedroom device set is currently unavailable for media playback (e.g., initiated from the device 800). FIG. 8E also depicts the text and image of indicator 810C different than the other device indicators, further visually indicating that the device set Bedroom is unavailable for media playback from device 800. For example, the device 800 can display an indicator for an unavailable device in a grey color, or other suitable manner that visually distinguishes the unavailable device from available devices.

[0346] At FIG. 8F, device 800 receives user input 814. In this example, user input 814 represents a contact (e.g., touch, tap, deep press, or press and hold) on a location outside of the displayed multi-device interface 810 (e.g., on a location that does not include display of an indicator of multi-device indicator 810). In some examples, in response to receiving user input 814, device 800 ceases displaying multi-device interface 810.

[0347] FIGS. 8G and 8H illustrate an exemplary current media interface. At FIG. 8F, device 800 displays current media interface 816. In some examples, current media interface 816 is displayed in response to user input 814 of FIG. 8F. In some examples, current media interface 816 is displayed in response to user input 812 of FIG. 8D. Current media interfaces 804 and 816 are similar, however current media interface 816 is an updated version of interface 804, and

reflects that the device 800 is now targeting a different media session (e.g., set of devices and/or media). In particular, current media interface 816 includes album art 816A (e.g., different than the album art 804A) media identifier 816B (song title "Heart Wave"), media identifier 816C (artist "MKTS"), media identifier 816D (album "MKTS"), and device selection affordance 816E. As shown in FIG. 8G, the device selection affordance 816E includes an identifier for another device set, "Kitchen + Living Room". Thus, in this example, the media item identified in the current media interface 816 is currently playing back on the device set "Kitchen + Living Room" (e.g., which excludes device 800). At FIG. 8H, device 800 receives a user input 818. In this example, user input 818 represents a contact (e.g., touch, tap, deep press, or press and hold) on a location outside of the displayed current media interface 816 (e.g., on a location that does not include display of an indicator of current media interface 816). In some examples, in response to receiving user input 818, device 800 ceases displaying current media interface 816. In some examples, device 800 displays browser interface 820 in response to user input 818.

[0348] FIGS. 8I and 8J illustrate an exemplary browser interface. At FIG. 8I, device 800 displays browser interface 820 for viewing and selecting a plurality of media items. Browser interface 820 depicts a music browsing interface, and in particular, an album view of a music album. Browser interface 820 includes identifier 820A identifying the currently-selected album, as well as a plurality of affordances (e.g., 820B and 820C), each corresponding to a media item (e.g., tracks of the music album). Current media indicator 822 identifies media currently playing back as part of a currently-selected media session (e.g., the track titled "Heart Wave" playing back on the devices Kitchen + Living Room). At FIG. 8J, device 800 receives user input 824 representing selection of media item affordance 820B. In some examples, user input is a tap or touch input at a location associated with affordance 820B. In some examples, user input is a deep press or touch and hold input at a location associated with affordance 820B.

[0349] FIG. 8K illustrates an exemplary current media interface. In some examples, device 800 displays current media interface 826 in response to user input 824 (e.g., shown in FIG. 8K). In some examples, device 800 displays current media interface 826 in response one or more user inputs received subsequent to user input 824. As shown in FIG. 8K, current media interface 826 illustrates that, in response to a request to selection of a media item for playback (e.g., selection of affordance 820B) and while device 800 is targeted to a set of devices (e.g., "Kitchen + Living

Room" as shown in FIG. 8H), that playback is initiated on the targeted set of devices. Thus, user input selection of affordance 820B associated with the music item titled "The Spiral" by artist "The Fomoers" causes display of current media interface 826 that includes corresponding appropriate indicators 826A-826D. Device selection affordance 826E indicates that the focus of device 800 continues to be the device set "Kitchen + Living Room" (e.g., as it was immediately prior to receiving user input 824).

[0350] FIGS. 8L-8P illustrate exemplary interfaces for initiating playback of media on one or more devices that are currently playing back other media. FIG. 8L depicts multi-device interface 810, as described above with respect to FIG. 8C. As shown in FIG. 8L, the currently-selected media session corresponds to the device iPhone and the media track "Same Thing".

[0351] At FIG. 8M, device 800 has received user input selection of the device set Kitchen (e.g., represented by the checkmark in indicator 810H of indicator 810D) and the device set Living Room (e.g., represented by the checkmark in indicator 810I of indicator 810E). In some examples, the user input selection of a device is user input 830 at a location associated with a device indicator (e.g., indicator 810D, as shown in FIG. 8M). FIG. 8M also depicts user input 832, representing selection of affordance 810L. In some examples, user input 832 is a contact (e.g., tap, touch, deep press) at a location associated with affordance 810L. In some examples, the user input selection of a device is user input 832. FIG. 8M also depicts indicator 810J, for a non-selected media session that includes playback on the devices Kitchen and Living Room. Accordingly, the device Kitchen and Living Room are currently busy (e.g., are part of a different media session (e.g., than the media session that is selected) that includes playback of media content) playing back the media item titled "Heart Wave". Further, FIG. 8M depicts indicator 810J with identifier 810M, which identifies a media session owner. In some embodiments, a media session owner is a device. In some embodiments, a media session owner is a user or (e.g., a user account). In some embodiments, a media session owner is a device and a user. As shown in FIG. 8M, the media session owner associated with the media session of indicator 810J includes both a user and a device, "Bob's iPhone". Thus, in this example, the media session owner includes a device different than device 800, and the different device is associated with a user (e.g., Bob) that is not associated with device 800 (e.g., device 800 is not logged into an account associated with Bob).

[0352] FIG. 8N illustrates an exemplary notification prompt for taking over playback of a device. In some embodiments, device 800 displays notification prompt 834 in response to receiving user input 830. In some embodiments, device 800 displays notification prompt 834 in response to receiving user input 832. In some examples, a notification prompt is displayed when device 800 attempts to initiate (or be configured to initiate) media playback on a device that is currently busy playing back media. In some examples, a device is busy playing back media if such media was initiated by a different media session owner. Thus, if device 800 is associated with a first user account (e.g., belonging to user Tim) and the targeted device is currently playing back media that was initiated by a second user account (e.g., from a device associate with a user account belonging to Bob) that is not associated with device 800, a notification prompt such as notification prompt 834 can be displayed.

[0353] In some examples, a notification prompt for taking over playback is displayed in response to user input 830. For example, in response to a user selecting an indicator for Kitchen device in FIG. 8M, a notification prompt is displayed. In some examples, a notification prompt for taking over playback is displayed in response to user input 832. For example, in response to receiving selection of affordance 810L in FIG. 8M (e.g., after receiving selection of the Kitchen and Living Room indicators 810D and 810E), notification prompt 834 is displayed. Notification prompt 834 includes a message 834A that reads: "The Kitchen and Living Room speakers are currently playing "Heart Wave". Are you sure you would like to play "Same Thing" in the Kitchen and Living Room?". Further, notification prompt 834 includes yes affordance 834B and cancel affordance 834C. In some examples, selection of affordance 834B causes device 800 to take over the playback of the selected devices. In some examples, selection of affordance 834C causes device 800 to forgo taking over the playback of the selected devices.

[0354] FIG. 8O illustrates an exemplary multi-device interface after taking over playback from another media session owner. At FIG. 8O, device 800 displays the multi-device interface 810 depicted in FIG. 8O in response to selection of yes affordance 834B, representing confirmation of that the device 800 should take over playback and add the Kitchen and Living Room devices to the currently-selected "Same Thing" media session. Notably, in FIG. 8O, the indicator 810J (as shown in FIG. 8L) is no longer displayed. In some examples, an indicator for a media session ceases to be displayed subsequent to its devices being taken over. In this

example, the media session that corresponded to indicator 810J (e.g., playback of "Heart Wave" on devices "Kitchen + Living Room") no longer exists (e.g., because each of its devices have been added to another media session). Thus, indicator 810J is not displayed in multi-device interface 810 of FIG. 8O.

[0355] FIG. 8P illustrates an exemplary multi-device interface after less than all devices of a media session have been taken over. FIG. 8P is similar to FIG. 8O except that the device set "Living Room" was not taken over (e.g., not selected at FIG. 8M) and added to the media session for the media item "Same Thing". Thus, the indicator 810J continues to be displayed in FIG. 8P, but has been updated to reflect that the Kitchen device is no longer a part of that media session. For example indicator 810J in FIG. 8P now identifies "Living Room" as the device set included in its corresponding media session, and no longer identifies "Kitchen + Living Room" (e.g., as in FIG. 8M).

[0356] FIG. 9 is a flow diagram illustrating a method for configuring an electronic device to initiate playback on a plurality of available playback devices using an electronic device in accordance with some embodiments. Method 900 is performed at a device (e.g., 100, 300, 500, 580) with a display. Some operations in method 900 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0357] As described below, method 900 provides an intuitive way for configuring an electronic device to initiate playback on a plurality of available playback devices. The method reduces the cognitive burden on a user for configuring an electronic device to initiate playback on a plurality of available playback devices, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to configure an electronic device to initiate playback on a plurality of available playback devices faster and more efficiently conserves power and increases the time between battery charges.

[0358] The electronic device (e.g., 800) displays (902), on the display (e.g., 802), a first current media interface (e.g., 804 of FIG. 8A) for the device, wherein the first current media interface includes: a representation (e.g., 804A, 804B, 804C, and/or 804D of FIG. 8A) of a first media item available for playback on the device (904), and a playback device selection affordance (906) (e.g., 804E). In some examples, a current media interface is any user interface

for browsing, viewing, or controlling media playback. For example, a current media interface is a "Now Playing" interface.

[0359] In some embodiments, the electronic device (e.g., 800) is currently playing the first media item. For example, the first current media interface is a "Now Playing" interface. In some embodiments, the representation of the first media item available for playback on the device is selectable (e.g., is an affordance), and can be selected to initiate playback of the first media item on the electronic device. For example, the first current media interface is a browsing interface that allows browsing of one or more media items (e.g., the first media item) while a media application on the electronic device (e.g., or the electronic device itself) is targeting the electronic device. In some embodiments, selection of a media item (e.g., the first media item) while the music application is targeted to a device causes playback of the media item to initiate on the targeted device in response.

[0360] In some embodiments, selection of a playback device selection affordance (e.g., 804E) allows one or more available playback devices to be selected. In some embodiments, the available playback devices include the electronic device. Selection of one or more available playback devices may be referred to as "targeting", "retargeting", "focusing", "refocusing" a media application or the electronic device, as described above. For example, a media application or device can be targeted to three discrete devices: the electronic device and two external speakers connected to the device via a wireless connection. In this example, while targeted to the three discrete devices in this example, a media item is selected for playback, and playback is initiated on the three discrete devices.

[0361] The electronic device (e.g., 800) receives (908) a first user input (e.g., 806 of FIG. 8B) representing selection of the playback device selection affordance (e.g., 804E).

[0362] In response to receiving the first user input (e.g., 806), the electronic device (e.g., 800) displays (910), on the display (e.g., 802), one or more affordances (e.g., affordances 804B-804K) associated with a plurality of available playback devices (e.g., the device sets iPhone, Bedroom, Kitchen, and/or Living Room of FIG. 8C) connected to the device (e.g., 800). In some examples, available playback devices are devices that are connected to the device such that the electronic device can control or otherwise use the available playback devices to initiate playback

of media content (e.g., via instructions transmitted over a communication link). In some examples, the available devices are connected (e.g., to the electronic device) via a wireless connection (e.g., via Wi-Fi, Apple Airplay, or the like). In some embodiments, the plurality of available playback devices includes the electronic device.

[0363] Displaying one or more affordances associated with a plurality of available playback devices connected to the device provides the user with visual feedback about the state of a plurality of devices that are available for media playback. Providing improved visual feedback to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0364] Displaying one or more affordances associated with a plurality of available playback devices connected to the device allows the user to collectively view a plurality of states for a plurality of available devices in one optimized interface. Providing an optimized interface to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0365] The electronic device (e.g., 800) receives (912) a second user input (e.g., 812 of FIG. 8D) representing selection of the one or more affordances (e.g., affordance 810J of FIG. 8D) associated with the plurality of available playback devices. In some examples, the electronic device receives selection of several devices (e.g., via selection of indicators 810H and 810I as shown in FIG. 8E), or receives selection of a single indicator (e.g., 810J FIG. 8D) representing a plurality of devices.

[0366] In response to receiving the second user input (e.g., 812), the electronic device (e.g., 800) configures (914) the device to initiate playback on the plurality of available playback devices (e.g., the devices in the set "Kitchen + Living Room" represented by affordance 810J of FIG. 8D, and individually by affordances 804D and 804E of FIG. 8E). In some embodiments, configuring the device to initiate playback comprises initiating playback on the plurality of

available playback devices. In some embodiments, playback is not initiated in response to configuring the device to initiate playback on the plurality of available playback devices, but if a command to initiate playback of media is subsequently received by the device, playback will be initiated on the plurality of available playback devices. Thus, in some examples, configuring a device to initiate media playback, on one or more devices, "retargets" the device (e.g., as described above) on the one or more devices, for example, to control and/or output playback on a plurality of available devices.

[0367] Configuring a device to initiate playback on the plurality of available playback devices allows the user to collectively control a plurality of available devices in one optimized interface. Providing an optimized interface to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0368] In some embodiments, configuring the device to initiate playback on the plurality of available playback devices comprises transmitting an instruction to the available playback devices to initiate playback (e.g., the devices in the set "Kitchen + Living Room" represented by affordance 810J of FIG. 8D, and individually by affordances 804D and 804E of FIG. 8E).

[0369] In some embodiments, while the electronic device (e.g., 800) is configured to initiate playback on the plurality of available playback devices (e.g., devices represented by affordances 804D and 804E of FIG. 8E), and after ceasing to display the one or more affordances (e.g., affordances 804B-804K) associated with the plurality of available playback devices (e.g., the device sets iPhone, Bedroom, Kitchen, and/or Living Room of FIG. 8C), the electronic device receives a third user input (e.g., 824 of FIG. 8J), representing a request to initiate playback of a second media item (e.g., represented by 820B of FIG. 8J). For example, while the electronic device is targeting the plurality of available playback devices, the third user input is received after ceasing to display multi-device interface 810. In response to receiving the third user input, the electronic device transmits a request to initiate playback of the second media on the plurality of available playback devices (e.g., the device set "Kitchen + Living Room" indicated by device

selection affordance 824E of FIG. 8K). In some embodiments, the third user input is received at a browser interface (e.g., 820 of FIG. 8I). In some examples, a browser interface is an interface for browsing and selecting media items for playback (e.g., a music application, a video application). The browser interface can be displayed after configuring the device to initiate playback on the plurality of available playback device at a multi-device interface, and after ceasing to display (e.g., dismissing, exiting) the multi-device interface. In some examples, the device remembers that playback of selected media will be initiated on the plurality of available devices that were selected at the multi-device interface (e.g., as shown in FIG. 8F). In some embodiments, the third user input is selection of an affordance of the browser interface. For example, such as an affordance associated with a media item, such as a song or video. In some embodiments, transmitting a request to initiate playback of the second media comprises transmitting an address (e.g., a URL) that is usable to stream the media. In some embodiments, transmitting a request to initiate playback of the second media comprises transmitting a stream of the media from the electronic device to the plurality of available devices.

[0370] In some embodiments, the first current media interface (e.g., 804 of FIG. 8A) for the electronic device (e.g., 800) includes an indication (e.g., 804A, 804B, 804C, and/or 804D) of a media item currently playing back on the device, and further in response to receiving the first user input (e.g., 806), the electronic device (e.g., 800) concurrently displays, on the display (e.g., 802): a first region (e.g., 828A of FIG. 8D) that includes the one or more affordances (e.g., 804B-804K) associated with the plurality of available playback devices; and a second region (e.g., 828B of FIG. 8D) that includes the indication (e.g., 804A, 804B, 804C, and/or 804D) of the media item currently playing back (e.g., "Same Thing" by "Jambug" as indicated by indicator 810A of FIG. 8D) on the electronic device (e.g., 800). Further in response to receiving the second user input (e.g., 812 of FIG. 8D), the electronic device (e.g., 800) visually updates the second region (e.g., 828B of FIG. 8D) to display an indication of a media item currently playing back (e.g., "Heart Wave" by "MKTS" as indicated by indicator 810J of FIG. 810J of FIG. 8E) on the plurality of available playback devices (e.g., the device set "Kitchen + Living Room" represented by indicator 810J of FIG. 8E).

[0371] In some embodiments, the one or more affordances (e.g., 810J of FIG. 8D) associated with the plurality of available playback devices connected to the device include information

(e.g., "Heart Wave" by "MKTS" as indicated by indicator 810J of FIG. 810J of FIG. 8D) identifying media currently being played back by the plurality of available playback devices.

[0372] In some embodiments, the information identifying the media currently being played back by the plurality of available playback devices includes one or more of: a song title, an artist, a movie title, an episode name, and a URL (e.g., "Heart Wave" by "MKTS" as indicated by indicator 810J of FIG. 810J of FIG. 8E).

[0373] In some embodiments, while the electronic device (e.g., 800) is configured to initiate playback on the plurality of available playback devices (e.g., the device set "Kitchen + Living Room" represented by indicator 810J of FIG. 8E), the electronic device transmits an instruction to initiate media playback of a third media item (e.g., "Same Thing" by "Jambug", as indicated by the active media session indicator 810A of FIG. 8M) on the plurality of available playback devices (e.g., the device set Kitchen and the device set Living Room respectively corresponding to indicators 810D and 810E of FIG. 8M), wherein the plurality of available playback devices are currently playing a fourth media item (e.g., "Heart Wave" by "MKTS", as indicated by the active media session indicator 810J of FIG. 8M) that was initiated by another device (e.g., the device named "Bob's iPhone" identified by indicator 810J of FIG. 8M) different than the electronic device (e.g., 800). Subsequent to transmitting the instruction to initiate playback of the third media item, the electronic device (e.g., 800) receives an indication (e.g., as shown by indicator 810A of FIG. 8O, checkmarks in indicators 810H and 810I) that the plurality of devices are currently playing back the third media item (e.g., "Same Thing" by "Jambug", as indicated by the active media session indicator 810A of FIG. 8O).

[0374] In some embodiments, the electronic device (e.g., 800) determines that the plurality of available playback devices (e.g., the device set Kitchen and the device set Living Room respectively corresponding to indicators 810D and 810E of FIG. 8M) are currently playing content (e.g., "Heart Wave" by "MKTS", as indicated by indicator 810J of FIG. 8M) that was initiated by another device (e.g., "Bob's iPhone" as indicated by indicator 810J of FIG. 8M). In accordance with the determination that the plurality of available playback devices are currently playing content that was initiated by another device, the electronic device requests (e.g., displays prompt 834 of FIG. 8N) user input confirmation (e.g., selection of affordance 834B or 834C of

FIG. 8N) of the instruction to initiate media playback of a third media item on the plurality of available playback devices. The electronic device (e.g., 800) receives a user input response (e.g., selection of affordance 834B or 834C of FIG. 8N) to the request for user input confirmation. In accordance with the user input response being a positive confirmation (e.g., selection of affordance 834B FIG. 8N), the electronic device transmits the instruction to initiate media playback of the third media item on the plurality of available playback devices. In accordance with the user input response being a negative confirmation (e.g., selection of affordance 834C FIG. 8N), forgoing transmitting the instruction to initiate media playback of the third media item on the plurality of available playback devices.

[0375] In some embodiments, further in response to receiving the second user input (e.g., 812 of FIG. 8D), the electronic device (e.g., 800) ceases displaying (e.g., replaces display with current media interface 816 of FIG. 8G) the one or more affordances (e.g., the affordances included in 810 of FIG. 8D) associated with the plurality of available playback devices.

[0376] In some embodiments, a media item includes one or more of audio data, video data, or both.

[0377] Note that details of the processes described above with respect to method 900 (e.g., FIG. 9 are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 900. For brevity, these details are not repeated below.

[0378] FIGS. 10A-10O illustrate exemplary user interfaces for configuring an electronic device, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 11.

[0379] FIG. 10A illustrates an exemplary smart speaker device. Device 1000 is a smart speaker device, and can include one or more features of device 580 as described above. In some embodiments, device 1000 includes one or more features of device 100, 300, or 500. As shown in FIG. 10A, device 1000 can include one or more LED indicators, which are shown displaying a rotating swirling pattern 1004 in FIG. 10A. In some examples, the LED indicators

display a pattern (e.g., also referred to as a "boot up pattern") when device 1000 is powered on for the first time (e.g., during boot up of the device). The pattern 1004 can be created by one or more LED indicators or other display elements (e.g., a display panel), which are referred to herein interchangeably, unless otherwise noted. In some examples, LED indicators are positioned under a touch-sensitive surface of device 1000. FIG. 10B illustrates three concentric rings, 1006A, 1006B, and 1006C that represent concentric rings of discrete LED indicators that comprise an LED indicator that forms pattern 1004. The rings 1006A-1006C can be used to create the boot up pattern (e.g., pattern 1004). In some examples, other arrangements of lighting elements are used. When device 1000 is finished booting up, or is otherwise ready to pair (e.g., with another device) and/or be configured, the pattern 1004, in some examples, can change. For example, the pattern may transition to a pulsing pattern (e.g., progressively lighting up each concentric ring 1006A-1006C, and then turning them off in reverse order).

[0380] As shown in FIG. 10A, device 1000 has limited display capabilities, and thus configuration of device 1000 can be made easier by the use of a second device having more robust display capabilities. In some examples, device 1000 is a smart home device (e.g., a network connected speaker) that a user places within their home (e.g., in one location, or moving between many locations). In such case, the user has physical access to the device when configuring it, and thus the use of physical proximity can be used as part of an intuitive and easy to use interface for configuring the device 1000. Thus, rather than making a user navigate various menus attempting to establish or access communication between a personal device (e.g., device 1010) and the device 1000, a user can place their personal device near device 1000, for example, as part of a configuration process.

[0381] FIGS. 10C-10G illustrate exemplary physical arrangements of a first device and a second device. FIG. 10C illustrates a first scenario that includes first device (e.g., device 1000) sitting on a table, and a user holding a second device (e.g., personal device 1010) at a long distance away from the first device. In some embodiments, device 1010 includes one or more features of device 100, 300, or 500. FIG. 10D illustrates an exemplary overhead view of the first scenario. As shown in FIG. 10D, device 1000 and device 1010 are far apart. FIG 10D includes the proximity condition range indicator 1008. A proximity condition range indicator is also referred to herein as a "proximity zone indicator" or simply as a "proximity zone". Device 1010

is not inside of proximity condition range indicator 1008. Proximity condition range indicator 1008 is included as a visual aid, and is intended to represent a physical proximity that would satisfy a proximity condition. For example, range indicator 1008 can represent the range of a near-field communication detection circuitry of device 1000. In some embodiments, any appropriate technique can be used to detect proximity between devices. For instance, in some examples, wide-band wireless connection is used. Wide-band wireless connection is used, for example, to determine one or more of: directionality, distance, and orientation, of one or more devices. Thus, presence of a detectable device within (e.g., partially or completely) within the proximity condition range indicator 1008 would satisfy a proximity condition, but would not if the detectable device is located outside of the range indicator 1008. One of skill would appreciate that a detection range of physical proximity can be non-uniform, can be affected by numerous variables (e.g., wireless interference, air humidity, or the like), and can include points in space in three dimensions, all of which are intended to be within the scope of this disclosure. Thus, the graphical representation of proximity condition range indicator 1008 is not intended to limit the scope of determining whether a proximity condition is satisfied.

[0382] FIG. 10E illustrates a second scenario that includes first device (device 1000) sitting on a table, and a user holding a second device (personal device 1010) at a short distance away from the first device. FIG. 10F illustrates an exemplary overhead view of the second scenario. As shown in FIG. 10F, device 1000 and device 1010 are close together and now device 1010 is at least partially within the proximity condition range indicator 1008. Because proximity condition range indicator 1008 represents a physical proximity that satisfies the proximity condition, the first device 1000 detects an indication that a physical proximity between the first device and the second device 1010 satisfies the proximity condition. In some examples, in accordance with detecting that a proximity condition is satisfied, device 1000 initiates a configuration process. For example, the device 1000 can enter a configuration mode for receiving configuration data (e.g., settings, account information, and the like) from the second device. In some examples, in accordance with detecting that a proximity condition is satisfied, device 1000 proceeds to a next step in a configuration process. For example, if the device 1000 is already in a configuration mode, the satisfaction of a proximity condition can cause the device 1000 to proceed to a next step. For instance, satisfaction of the proximity condition can be required in order to verify the identity of the second device or verify that the user of the second

device has physical access to device 1000 (e.g., is not a user outside of the home that is attempting to remotely access and configure device 1000).

[0383] FIGS. 10G-10J illustrate exemplary techniques for using audio tone signals when configuring a device. FIG. 10G depicts an overhead view similar to FIG. 10F, but additionally depicts device 1000 outputting an audio tone signal 1014 (e.g., depicted as graphical representations of sound waves). In this example, speaker 1000 plays an audible tone. In some examples, the tone is part of a proximity handshake tone at the beginning of a configuration process (e.g., to initiate the process), or a tone later in (e.g., during) the configuration process. In some embodiments, the audio tone signal is out-of-band to a communication link between the first device and the second device. For example, the first and second devices are connected and exchange data communications over a Bluetooth wireless connection (e.g., a communication link). In this example, while the devices continue to be connected via the Bluetooth connection (e.g., the devices detect each other of Bluetooth and perform a handshake), an audio tone signal (e.g., an audible tone that includes or represents a communication between the devices) is out-of-band with the Bluetooth communication link. In some embodiments, being out-of-band to a communication link means that a signal (e.g., an audio tone signal) is not communicated using the same data exchange technique. In some examples, using a different data exchange technique can include one or more of using a different: transmission protocol (e.g., Bluetooth, Wi-Fi), data carrier technology (e.g., sound (mechanical) waves instead of light (electromagnetic) waves), hardware (e.g., wireless antenna instead of a wired connector), or the like. In some embodiments, an audio tone signal is a sound traveling through a transmission medium (e.g., that is detectable by another device). For example, an audio tone signal can include frequencies inside of the human audible frequency range (e.g., approximately 20 Hertz to 20,000 Hertz), and/or include frequencies outside of the human audible frequency range.

[0384] FIG. 10H illustrates an exemplary perspective view of the overhead view of FIG. 10G. As can be seen, a user is holding second device 1010 a short distance away from device 1000, which is outputting the audio tone signal 1014.

[0385] FIG. 10I illustrates an exemplary perspective view where the second device 1010 is outputting a second audio tone signal 1016 (e.g., which may referred to as a response audio tone

signal). In some embodiments, a response audio tone signal 1016 (also referred to as a "second audio tone signal") is output by the second device in response to the second device detecting the audio tone signal by device 1000. For example, device 1010 can play a response tone immediately in response to detecting the tone signal 1014, in order to communicate to the device 1000 that the first audio tone signal was received. In some embodiments, a response tone signal is a tone signal that is output by the second device at a later step in the configuration process. For example, the first tone signal (e.g., 1014) can represent an audible indication that the configuration process has been initiated (e.g., and proceeds to a first step of the process), and subsequent to initiating the configuration process for the second device (e.g., using the first device), the second device can output the second audio tone signal (e.g., 1016) at a fourth step of the configuration process. In some embodiments, the response tone signal includes encoded data. For example, as part of the configuration process for the second device, the second device (e.g., a personal device 1010 that is already associated with a user's account) can pass account login information to the first device (e.g., so that the device 1000 can access or update data associated with the user's account) via encoded data in an audio tone signal. This technique can be used, for example, when transmission of such data by other means may be unsecure (e.g., may be intercepted). In some examples, data is encoded in an audio tone signal by any appropriate technique, as one of skill in the art would appreciate. For example, an audio tone signal can include tones at a particular frequency representing data signatures. In some examples, an audio tone signal includes modulated carrier sound waves that carry data information.

[0386] In some embodiments, the first device repeats the audio tone signal until it receives a confirmation that the second device detected the audio tone signal. For example, device 1000 can repeat output of audio tone signal 1014 until receiving confirmation that device 1010 detected the signal 1014. For example, the confirmation can be a response tone (e.g., second audio tone signal 1016; "out-of-band") or a communication over the communication link (e.g., Bluetooth, Wi-Fi, or the like; "in-band") between the two devices.

[0387] In some embodiments, the first device repeats the audio tone signal for a predetermined number of times. In some embodiments, the first device repeats the audio tone signal for a predetermined number of times or until a confirmation that the second device detected the audio tone signal is received, whichever comes first. For example, device 1000 can

repeat the audio tone signal 1014 three times before it ceases repeating the tone, if no confirmation is received (e.g., response tone signal 1016).

[0388] In some embodiments, the first device outputs a dictated audio passcode. As shown in FIG. 10J, device 1000 outputs an audio tone signal 1018 that includes the dictated text "The passcode is 061917". In this example, the user can enter the passcode "061917" into device 1010, for example, to initiate or proceed to a next step in the configuration process for configuring the first device 1000. Thus, if the audio tone signal is not detected by the second device (e.g., device 1000 does not receive a conformation), device 1000 can output a dictation of a passcode that can, for example, be used to proceed to a next step in the configuration process or authorize data exchange between the two devices. For example, a user can enter the dictated passcode into device 1010, or repeat the passcode back to device 1000 (e.g., using voice input). Thus, if the audio tone signal was being used as part of a process of confirming the identity of the first device or second device, but the second device cannot detect the audio tone signal, the user can use the dictated audio passcode to achieve the same effect. For example, using the passcode confirms that the device 1010 is within a relatively close physical proximity by virtue of the user hearing and entering the dictated audio passcode.

[0389] In some embodiments, device 1000 outputs encoded data visually. For example, LED indicators (e.g., 1006A-1006B) can be used to output a pattern or sequence of light emission that is detectable by another device (e.g., 1010) and that includes or otherwise represents data information. The other device can, for example, capture or record the light pattern and decode the data, or transmit the captured data back to the device to verify that the other device observed the light pattern.

[0390] In some embodiments, during a configuration process, configuration data is received and stored by device 1000. Examples of configuration data include, but are not limited to, one or more of: whether a device (e.g., 1000 or 1010) is a member of a group of devices (e.g., home devices, or a stereo pair), a name or identifier of the device, a user associated with the device, a location (e.g., a room name) of the device, user account credentials (e.g., for network or cloud-based services such as iCloud, iTunes, Apple Music, Spotify, or the like). Configuration data for

device 1000 can also be added or edited in a home control application (e.g., as shown in FIG. 6A).

[0391] In some embodiments, the first device is configured to be included in a group of devices. FIG. 10K illustrates two devices—first device 1000 and third device 1020—that make up an exemplary group of devices. As shown in FIG. 10K, the first and third devices are both smart speaker devices (e.g., such as device 580). In some embodiments, each device in the group of devices is configured to output an audio channel. FIG. 10K shows the first and third devices placed on either side of a television. In this example, device 1000 is configured to output a left audio channel (e.g., of a stereo audio signal) and the device 1020 is configured to output a right audio channel (e.g., of a stereo audio signal). Thus, the devices in FIG. 10K can be referred to as a "stereo pair", a "stereo group", or the like. In some embodiments, devices in a group are configured to output identical channels. For example, referring gain to FIG. 10K, both devices 1000 and 1020 can be configured to output the same audio signal (e.g., monaural audio). In some examples, a group of devices includes more than two devices. For example, FIG. 10L illustrates a group of devices comprised of device 1000, device 1020, device 1022, and device 1024. As described with reference to FIG. 10K, each device in the group can be configured to output a different audio channel signal (e.g., "rear left" signal on device 1022, "rear right" signal on device 1024, "front left" signal on device 1000, and "front right" signal on device 1020).

[0392] In some embodiments, a group of devices (e.g., as shown in FIGS. 10K and 10L), are treated as one device at a user interface for selecting devices for playback. For example, the group of devices 1000 and 1020 of FIG. 10K can be configured as a device set named "Living Room". In some examples, a multi-device interface (e.g., multi-device interface 610 of FIG. 6J, as described above) would display one individual indicator (e.g., indicator 610N of FIG. 6J) corresponding to the stereo pair of devices 1000 and 1020. In some examples, the individual indicator includes an indication that it represents a group of devices (e.g., indicator 610N of FIG. 6J can include two graphical speaker icons).

[0393] In some embodiments, a group of devices (e.g., as shown in FIGS. 10K and 10L) are displayed separately at a user interface for selecting devices for playback. In some examples, a multi-device interface (e.g., multi-device interface 610 of FIG. 6J, as described above) includes

an individual indicator for each device in a group of devices. For example, referring back to the group in FIG. 10K, device 1000 can be made to output a first media item while device 1020 outputs a second media item, concurrently (e.g., using a multi-device interface such as 610 of FIG. 6J).

[0394] FIG. 10M depicts an exemplary simplified network diagram of a device group, in accordance with some embodiments. In some embodiments, the first device and the third device establish a communication link between, and communicate with, each other directly. For example, FIG. 10M depicts an access point 1030, which may represent a wireless router, network switch, or the like, that creates a location area network that is connected to both devices 1000 and 1020. In this example, device 1000 is connected to access point 1030 (e.g., as illustrated by communication link 1034). In this example, device 1020 is also connected to access point 1030 (e.g., as illustrated by communication link 1036). In this example, access point 1030 is a wireless router and devices 1000 and 1020 are connected to the same wireless (e.g., Wi-Fi) network generated by the access point 1030. Access point 1030 is also connected to second device 1010 (e.g., illustrated by communication link 1038) as well as to a wide area network (e.g., internet). Thus, device 1000 and 1020 are connected by access point 1030. Devices 1000 and 1020, in some examples, can communicate directly, bypassing the need to exchange communications via access point 1030. For example, devices 1000 and 1020 can create a Wi-Fi network between them (e.g., where one device acts as an access point for the other). Communication directly with each other, rather than through an access point that each device in the group is connected to, can provide useful benefits. For example, if both devices 1000 and 1020 have a poor connection to access point 1030, synchronization of simultaneous audio playback on the group of devices can be degraded due to increased latency. If devices 1000 and 1020 have a strong connection directly between them, then latency issues are reduced. Further, routing communication directly between devices 1000 and 1020 reduces usage of the access point 1030's bandwidth by reducing traffic on the network created by the access point (e.g., the user's main home Wi-Fi network). As described previously, in some examples, devices 1000 and 1020 do not have to be grouped to communicate directly. For instance, in some examples, devices 1000 and 1020 can communicate directly out-of-the-box, without being configured by the user to form a device group.

[0395] FIG. 10M also depicts communication link 1039 between device 1010 and device 1000. In some examples, link 1039 is a Bluetooth connection, a Wi-Fi connection, or the like.

[0396] FIGS. 10N-10O illustrated exemplary techniques for indicating progress in a configuration process. In some examples, the first device and the second device output synchronized audio tone signals. The synchronized audio tone signals, for example, can be used to indicate progression through a configuration process. In some examples, the synchronized tones are output at the completion of each step of a configuration process. For example, FIG. 10N illustrates device 1000 outputting audio tone signal 1042, and device 1010 is outputting audio tone signal 1044. Audio tone signals 1042 and 1044 are outputting by their respective devices to be synchronized in time (e.g., output such that at least a portion of each tone are played during an overlapping period of time).

[0397] In some embodiments, synchronized audio tone signals are harmonics. Thus, in some examples, the synchronized audio tone signals are the same audio frequency. For example, audio tone signal 1042 has a frequency of 2000 Hertz, while audio tone signal 1044 has a frequency of 2000 Hertz. In some examples, audio frequency of one of the synchronized audio tone signals is an integer multiple of the audio frequency of the other frequency. For example, audio tone signal 1042 has a frequency of 1000 Hertz, while audio tone signal 1044 has a frequency of 2000 Hertz.

[0398] In some embodiments, synchronized audio tone signals are each a different frequency of a musical chord or a musical scale. Thus, for example, the synchronized tones sound (e.g., to a user) like they are in harmony.

[0399] In some embodiments, subsequent synchronized audio tone signals have a different audio frequency. For example, to indicate that the configuration process is progressing, the synchronized audio tone signals can increase (e.g., step up) in frequency at each subsequent step in a configuration process. In the example shown in FIG. 10N, synchronized audio tone signals 1042 and 1044 are output by their respective devices upon completion of a first step in a configuration process.

[0400] Turning now to FIG. 10O, device 1000 outputs synchronized audio tone signal 1046 and device 1010 outputs synchronized audio tone signal 1048. In this example, the signals 1046 and 1048 are output subsequent to signals 1042 and 1044, and are output upon completion of a second step in the configuration process. Further, signals 1046 and 1048 can be harmonics (e.g., the same audio frequency or integer multiples of the same frequency) or compatible tones (e.g., frequencies in the same musical chord or musical scale). To illustrate the concept that the synchronized audio tone signals change in frequency as a configuration process progress, plot 1040 of FIG. 10O is helpful. Plot 1040 depicts a first point 1040A at a frequency of 2000 Hertz and a second point 1040B at a frequency of 8000 Hertz. First point 1040A represents the frequency of audio tone signals 1042 and 1044, which are output after the first step in the configuration process. Second point 1040B represents the frequency of audio tone signals 1046 and 1048, which are output after the subsequent second step in the configuration process. In some examples, synchronized audio tone signals decrease in frequency as progress during the configuration process progresses. Thus, using synchronized audio tone signals, the first and second devices can audibly indicate progress of a configuration process to the user. For example, the audio tone signals can give the impression of approaching a crescendo or a final completion frequency. As one of skill would appreciate, synchronized audio tone signals can be used to indicate progress of other processes involving two or more devices. Further, as should be apparent, more than two devices can be configured to output synchronized audio tone signals.

[0401] FIG. 11 is a flow diagram illustrating a method for configuring an electronic device using a second device in accordance with some embodiments. Method 1100 is performed at a device (e.g., 100, 300, 500, 580). Some operations in method 1100 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0402] As described below, method 1100 provides an intuitive way for configuring a first device using a second device. The method reduces the cognitive burden on a user for configuring a first device using a second device, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to configure a first device using a second device faster and more efficiently conserves power and increases the time between battery charges.

[0403] The electronic device (e.g., 1000) detects (1102) an indication that a physical proximity between the electronic device (e.g., 1000 of FIG. 10F) and a second device (e.g.,1010 of FIG. 10F) satisfies a proximity condition (e.g., device 1010 is within proximity zone 1008, as illustrated in FIG. 10F), wherein the electronic device (e.g., 1000) is connected to the second device (e.g., 1010) via a communication link (e.g., 1039). In some examples, the electronic device and second device are connected via a wireless data transmission protocol (e.g., Bluetooth, Wi-Fi, or the like). In some embodiments, the indication is detected using a near-field communication technique.

[0404] In accordance with detecting the indication, the electronic device audibly outputs (1104) an audio tone signal (e.g., 1014), wherein the audio tone signal is out-of-band to the communication link (e.g., 1039). For example, the device is a speaker that plays a tone. In some examples the tone is a proximity handshake tone at the beginning of a configuration process. In some examples the tone is played by the speaker during the configuration process (e.g., after it begins). In some embodiments, being out-of-band to a communication link means that a signal (e.g., an audio tone signal) is not communicated using the same data exchange technique.

[0405] Using a detection of physical proximity allows the user to efficiently move through a configuration process for a device with fewer required user inputs. Providing an optimized configuration process to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0406] Subsequent to outputting the audio tone signal (e.g., 1014), the electronic device (e.g., 1000) determines (1106) whether a response audio tone signal (e.g., 1016) from the second device (e.g., 1010) has been received.

[0407] In accordance with a determination that the response audio tone signal (e.g., 1016) from the second device (e.g., 1010) has been received, the electronic device (e.g., 1000) proceeds (1108) to a next step in a configuration process. In some embodiments, a configuration process includes receiving one or more configuration settings for the electronic device. In some

examples, configuration settings include one or more of: a location (e.g., room name) of the electronic device, a device identifier (e.g., a name), whether the electronic device is a member of a group of devices (e.g., a left or right channel of a stereo pair of speakers), user account credentials (e.g., allowing the electronic device to log into a cloud-based service), audio output preferences (e.g., language, volume), content restrictions (e.g., parental controls), or the like.

[0408] Proceeding to a next step in a configuration process in accordance with a determination that a response audio tone signal has been received allows the user to efficiently move through a configuration process for a device with fewer required user inputs. Providing an optimized configuration process to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0409] In accordance with a determination that the response audio tone signal from the second device has not been received, the electronic device (e.g., 1000) forgoes (1110) proceeding to the next step in the configuration process. For example, if the device does not detect a response audio tone signal (e.g., 1016) within a threshold amount of time after outputting an audio tone signal, the device can timeout (e.g., stop listening for a response audio tone signal), can repeat the tone, or dictate an audio passcode aloud.

[0410] In some embodiments, the response audio tone signal (e.g., 1016) represents an indication that the audio tone signal (e.g., 1014) was detected by the second device (e.g., 1010).

[0411] In some embodiments, at least one of the audio tone signal (e.g., 1014) and the response audio tone signal (e.g., 1016) comprises encoded data. In some embodiments, the encoded data comprises an indication of the outputting device's identity. For example, the encoded data can include an audible representation of data sent from one device to the other. Audibly outputting the data from one device such that the other device can detect it (e.g., using a microphone) can provide sufficient indication that the second device and the electronic device are within a relative close physical proximity to each other. In some embodiments, the encoded data includes user account credentials. For example, the encoded data can include data

representing user account information for a cloud-based service (e.g., iCloud login information), a media streaming service (e.g., Apple Music), or data for confirming that the electronic device (e.g., 1000) should be provided access to the user's account or personal data.

[0412] In some embodiments, the response audio tone signal (e.g., 1016) comprises encoded data, and further in accordance with a determination that the response audio tone signal from the second device (e.g., 1010) has been received, the electronic device (e.g., 1000) converts the encoded data, of the response audio tone signal, into decoded data. In some embodiments, the electronic device (e.g., 1000) uses at least a portion of the decoded data to log in to a user account of a cloud-based service. For example, if the encoded data includes login credentials for a cloud-based service (e.g., iCloud) the electronic device decodes the encoded data into an account user name and/or password, and logs into the cloud-based service.

[0413] In some embodiments, the response audio tone signal (e.g., 1016) comprises data that confirms the identity of the second device (e.g., 1010). For example, the data that is encoded can be data that was transmitted to the second device that is played back as a tone, or can be a unique identifier associated with the second device.

[0414] In some embodiments, the electronic device (e.g., 1000) includes a memory, and the electronic device receives one or more configuration settings from the second device (e.g., 1010). In response to receiving the one or more configuration settings, the electronic device stores the configuration settings in the memory of the electronic device. In some examples, configuration settings include one or more of the settings as described above.

[0415] In some embodiments, the one or more configuration settings includes a location identifier representing a physical location of the electronic device (e.g., 1000). For example, a location identifier can include a room identifier such as "Living Room", "Bedroom", "Bedroom 2", "Kitchen", "Tim's Room", or the like. In some embodiments, the location identifier is associated with the electronic device (e.g., 1000) and communicated to other devices (e.g., 1010). For example, the identifier can be used in a multi-device interface (e.g., 610 of FIG. 6J) or in a device management application (e.g., as shown in FIG. 6A) to provide an intuitive and/or descriptive identification of the device.

[0416] In some embodiments, the one or more configuration settings includes an indication that the electronic device (e.g., 1000) is included in a device group (e.g., a group comprised of device 1000 and 1020 of FIG. 10K), wherein the device group includes the electronic device (e.g., 1000 of FIG. 10K) and a third device (e.g., 1020 of FIG. 10K). For example, a device group includes two or more devices (e.g., smart speakers) configured to coordinate (e.g., synchronize) output of media (e.g., audio). In some embodiments, the devices in a device group (e.g., a speaker pair, such as left and right channels 1000 and 1020 of FIG. 10K, respectively) are treated as one unit. For example, the devices in a device group can be identified (e.g., at a multi-device user interface) by a single device group identifier (e.g., "Living Room") rather than as individual devices. Thus, a command to "Play audio in the Living Room" can cause all devices in the group "Living Room" to playback the audio.

[0417] In some embodiments, the first device (e.g., 1000) and the second device (e.g., 1010) are connected to a local area network via an access point (e.g., 1030). In some embodiments, the electronic device (e.g., 1000) establishes a communication link (e.g., 1032) to the third device (e.g., 1020 of FIG. 10M), wherein the communication link to the third device does not include the access point (e.g., 1030). In some embodiments, the communication link utilizes a wireless communication protocol. In some embodiments, the communication link is a wired connection between the electronic device and the third device. In some examples, the access point is a network router, a network switch, or other hardware for communicatively coupling a plurality of devices. In some embodiments, the electronic device (e.g., 1000) synchronizes media playback with the third device (e.g., 1010) using the communication link (e.g., 1032) to the third device.

[0418] In some embodiments, to audibly output the audio tone signal, the electronic device (e.g., 1000) audibly outputs the audio tone signal (e.g., 1014), and determines whether the second device (e.g., 1010) detected the audio tone signal. In some embodiments, determining whether the second device detected the audio tone signal includes determining whether data is received over the communication link (e.g., 1038) representing confirmation that the second device (e.g., 1010) detected the audio tone signal (e.g., 1014). In some embodiments, the electronic device receives (e.g., detects) a response audio tone signal (e.g., 1016) from the second device (e.g., 1010) representing confirmation that the second device detected the audio tone signal (e.g., 1014). In accordance with a determination that the second device (e.g., 1010)

did not detect the audio tone signal (e.g., 1014), the electronic device (e.g., 1000) repeats audibly outputting the audio tone signal (e.g., 1014). For example, the electronic device plays the tone (e.g., 1014) a second time (or an *nth* time). In accordance with a determination that the second device detected the audio tone signal, the electronic device forgoes repeating audibly outputting the audio tone signal.

[0419] In some embodiments, further in accordance with a determination that the second device (e.g., 1010) did not detect the audio tone signal (e.g., 1014), the electronic device (e.g., 1000) determines whether the audio tone signal has been audibly outputted a threshold number of times. In accordance with a determination that the audio tone signal has been audibly outputted a threshold number of times, the electronic device (e.g., 1000) audibly outputs a dictated passcode (e.g., 1018 of FIG. 10J). For example, the threshold number of times can be three. Thus, if the electronic device plays a tone (e.g., 1014) once and repeats the tone twice thereafter, without receiving confirmation (e.g., 1016) that the second device detected the audio tone signal, the electronic device can (e.g., on the fourth occurrence) output a dictated passcode instead of replaying the tone. In some embodiments, a passcode includes one or more alphanumeric characters. In some embodiments, audibly outputting a dictated passcode includes reading out individual characters of the passcode. For example, a text-to-speech algorithm is used to audibly output the passcode (e.g., as a phrase, as individual characters). In some embodiments, audibly outputting a dictated passcode includes reading out a phrase (e.g., formed by a plurality of alphanumeric characters). In accordance with a determination that the audio tone signal has not been audibly outputted a threshold number of times, the electronic device forgoes audibly outputting a dictated passcode. In some embodiments, the electronic device repeats outputting the audio tone signal until reaching the threshold number of times.

[0420] In some embodiments, the audio tone signal (e.g., 1014) is a first audio tone signal, and in response to progressing from a first step in the configuration process to a second step in the configuration process, the electronic device (e.g., 1000) outputs a second audio tone signal (e.g., 1042 of FIG. 10N), wherein the second audio tone signal is outputted concurrently with the output of a third audio tone signal (e.g., 1044 of FIG. 10N) outputted by the second device (e.g., 1010), and wherein the second audio tone signal and the third audio tone signal are harmonics. For example, the second audio tone signal and the third audio tone signal are the same frequency

or are integer multiples of the same frequency. In some examples, concurrent output of harmonic frequencies by the electronic device and the second device can provide audible feedback to a user of progress through the configuration process. For example, higher pitched (e.g., higher frequency) tones can indicate that the configuration process is near completion. Thus, relative or absolute progress through the configuration process can be provided in an intuitive manner.

[0421] In some embodiments, in response to progressing from the second step in the configuration process to a third step in the configuration process, the electronic device (e.g., 1000) outputs a fourth audio tone signal (e.g., 1046 of FIG. 10O), wherein the fourth audio tone signal is outputted concurrently with the output of a fifth audio tone signal (e.g., 1048 of FIG. 10O) outputted by the second device (e.g., 1010), wherein the fourth audio tone signal and the fifth audio tone signal are harmonics, and wherein the fourth audio tone signal is a higher frequency tone than the second audio tone signal (e.g., as illustrated by 1040). In some examples, the concurrently played tones escalate in frequency as progression through the configuration process increases.

[0422] Note that details of the processes described above with respect to method 1100 (e.g., FIG. 11) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 1100. For brevity, these details are not repeated below.

[0423] FIGS. 12A-12AM illustrate exemplary user interfaces for controlling media playback, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 13.

[0424] As described above (e.g., with respect to FIGS 10A-10F), devices can use the satisfaction of a proximity condition to perform one or more functions. The use of proximity of one device to another device can be used as a clear indicator that a user (e.g., holding one of the devices) would like to perform some action (e.g., invoke an interface) on one or both of the devices. For example, this can prevent a waste of device resources by avoiding excessive user input (e.g., to navigate one or more menus on a device display) in order perform a function.

Moreover, this can save the user time as well, for example, by reducing the number of user inputs required to perform a function (e.g., invoke an interface on a display).

[0425] In some examples, the satisfaction of a proximity condition between two devices can be used as an indication that a user (e.g., of one of the devices, a mobile device) would like to transfer media playback to or from one device (e.g., the mobile device) to the other device (e.g., a stationary smart speaker). Exemplary techniques are described below.

[0426] FIGS. 12A-12F illustrate an exemplary scenario in which a first device and a second device are not playing back media content when they are placed in close proximity.

[0427] FIG. 12A illustrates an exemplary scenario in which a user is holding device 1200 (e.g., a personal device, such as device 100, 300, 500) in close proximity to device 1202 (e.g., a smart speaker, such as device 580). In some embodiments, device 1200 includes one or more features of devices 100, 300, or 500. Thus, FIG. 12A is similar to FIG. 10E, the description of which is hereby incorporated by reference. In the example depicted in FIG. 12A, device 1200 and device 1202 are both not currently outputting audio (e.g., are not playing back media). In the example depicted in FIG. 12A, the physical distance between device 1200 and device 1202 satisfies a proximity condition. In some examples, device 1200 detects an indication that the proximity condition is satisfied (e.g., and, in response, initiates communication with device 1202, for example, to send an indication that the condition is satisfied). In some examples, device 1202 detects an indication that the proximity condition is satisfied (e.g., and, in response, initiates communication with device 1200, for example, to send an indication that the condition is satisfied).

[0428] It is noted that FIG. 12A, and the figures that follow, are not necessarily to scale and are included merely as a visual aid. Thus, unless otherwise noted, they are not intended as a limitation on a distance required to be in close proximity or to satisfy a proximity condition.

[0429] FIG. 12B illustrates an exemplary media information interface. FIG. 12B, also depicts an overhead view of the scenario of FIG. 12A. As can be seen, device 1200 is within (e.g., at least a portion of the device inside of) the proximity zone 1204 (also referred to as proximity condition range indicator 1204). Device 1200 is inside of proximity condition range

indicator 1204, and thus the proximity condition is satisfied. The phrase "close proximity" or "in close proximity", as used herein, refers to a scenario in which a physical distance between two devices satisfies a proximity condition (e.g., as shown in FIG. 12B and FIG. 10F). Proximity condition range indicator 1204 is included as a visual aid, and is intended to represent a physical proximity that would satisfy a proximity condition. For example, range indicator 1204 can represent the range of a near-field communication detection circuitry of device 1202. Thus, presence of a detectable device within (e.g., partially or completely) the proximity condition range indicator 1204 would satisfy a proximity condition, but would not if the detectable device is located outside of the range indicator 1204. One of skill would appreciate that a detection range of physical proximity can be non-uniform, can be affected by numerous variables (e.g., wireless interference, air humidity, device orientation, among others), and can include points in space in three dimensions, all of which are intended to be within the scope of this disclosure. Thus, the proximity condition range indicator 1204 is not intended as a limitation on the scope of determining whether a proximity condition is satisfied.

[0430] FIG. 12B illustrates media information interface 1208 displayed on a display 1206 of device 1200. In some embodiments, device 1200 includes one or more features of device 100, 300, or 500. In some examples, media information interface 1208 is displayed in response to device 1200 being placed (e.g., briefly or continuously) in close proximity with device 1202. Thus, because media information interface 1208 displayed in response to the device being placed in close proximity, it can referred to as a type of "proximity card" or specifically a "media proximity card".

[0431] In the example shown in FIG. 12B, device 1202 is a device associated with the location (or name) "Kitchen" and is a speaker (e.g., a smart speaker). Thus, proximity card 1208 includes the title 1208B "Kitchen Speaker", visually indicating the name of the speaker that the user placed device 1200 in close proximity to. Proximity card 1208 also includes an album affordance 1208A representing media content, as well as an identifier 1208C describing the status of the media content. In this example, the media content displayed in proximity card 1208 is content that was recently played on the second device, device 1202 (e.g., Kitchen Speaker). Thus, in this example, when neither device 1200 or device 1202 are currently playing back media content, placing the devices in close proximity causes device 1200 to display an interface

for viewing recently played media. In some examples, the recently played media was recently played by the first device (e.g., device 1200). In some examples, the recently played media is media that was recently played on both devices 1200 and 1202. In some embodiments, proximity card 1208 is displayed when one or more of devices 1200 and 1202 are currently playing back media.

[0432] At FIG. 12C, device 1200 receives user input 1210 on album affordance 1208A. In some examples, user input 1210 is a tap, touch, deep press, or a press and hold input received at device 1200. In some examples, a first device (e.g., device 1200) continues displaying a proximity card after the first device is no longer in close proximity to a second device (e.g., device 1202).

[0433] At FIG. 12D, in response to receiving user input 1210, device 1200 displays current media interface 1212. Current media interface 1212 (also referred to as "now playing screen" 1212) depicts one or more media items that are currently playing back (e.g., on the device 1202). Interface 1212 includes an identifier 1212A of a media item (e.g., the song titled "Same Thing" by artist "Jambug" from the album "Fly Away"), album art 1212B, and device selection affordance 1212C. As described above, a device selection affordance can indicate the current focus of the device. In this example, FIG. 12D depicts that the current media interface 1212 is targeting the Kitchen Speaker. Thus, in response to the user input selection of the affordance 1208A (e.g., representing a media item), device 1200 transmits an instruction to device 1202 to initiate playback of the corresponding media item, and updates the display of device 1200 to include playback controls and a media status of the device 1202. Accordingly, the user is provided a convenient technique for quickly beginning and controlling media playback on a second device (e.g., Kitchen Speaker) using a first device (e.g., person device 1200).

[0434] In some examples, current media interface is displayed on lock screen of device 1200. For example, the proximity card 1208 can be displayed while the device 1200 was locked and, while remaining locked, the device receives user input that causes the device to initiate playback of media and update the lock screen (e.g., to display interface 1212).

[0435] FIGS. 12E-12F illustrate exemplary techniques for browsing media within a media information interface, in accordance with some embodiments. At FIG. 12E, device 1200

receives user input 1214 on affordance 1208A, wherein affordance 1208B is visible. FIG. 12F shows user input 1214 shift to the left. In this example, user input 1214 was a leftward swipe on the recently played media affordances of proximity card 1208. In some embodiments, user input 1214 is a directional swipe in any direction (e.g., up, down, left, right, etc.). As shown in FIG. 12F, the affordances representing media items have shifted in accordance with the user input 1214. Affordance 1208B is now centered within proximity card, and affordance 1208C (e.g., representing a third media) is now visible.

[0436] FIGS. 12G-12N illustrate an exemplary scenario in which a first device and a second device are both playing back media content when they are placed in close proximity.

[0437] FIG. 12G illustrates an exemplary scenario in which a user is holding device 1200 (e.g., a personal device) in close proximity to device 1202 (e.g., a smart speaker). Thus, FIG. 12G is similar to FIG. 12A, the description of which is hereby incorporated by reference. In the example depicted in FIG. 12G, device 1200 and device 1202 are both currently outputting audio, as represented by sound waves 1216 (from device 1200) and sound waves 1218 (from device 1202).

[0438] FIG. 12H illustrates the states of devices 1200 and 1202 prior to them being placed in close proximity. As can be seen, device 1200 is currently playing back a first media item, as presented by current media interface 1220. Media indicator 1220A identifies the song titled "Monday Morning" currently playing on device 1200, album art 1220B, device selection affordance 1220C (indicating that device 1200 is targeted on itself, the device named "iPhone"), and playback control affordances 1220D. Device 1202 is currently playing back a media item as well, the song titled "Same Thing" by artist "Jambug", as represented by media state 1222.

[0439] FIG. 12I illustrates an exemplary media information interface. FIG. 12I, also depicts an overhead view of the scenario of FIG. 12G. As can be seen, device 1200 is within (e.g., at least a portion of the device is inside) the proximity zone 1204 (also referred to as proximity condition range indicator 1204). Device 1200 is inside of proximity condition range indicator 1204, and thus is in close proximity (e.g., the proximity condition is satisfied).

[0440] FIG. 12I illustrates media information interface 1224 displayed on a display 1206 of device 1200. In some examples, media information interface 1224 is displayed in response to device 1200 being placed (e.g., briefly or continuously) in close proximity with device 1202. Thus, because media information interface 1224 displayed in response to the device being placed in close proximity, it is also referred to as a "proximity card" or specifically a "media proximity card".

[0441] In the example shown in FIG. 12I, device 1202 is a device associated with the location (or name) "Kitchen" and is a speaker (e.g., a smart speaker). Thus, proximity card 1224 includes the title 1224B "Kitchen Speaker", visually indicating the name of the speaker that the user placed device 1200 in close proximity to. Proximity card 1224 also includes album affordances 1224A and 1224D representing media content (e.g., media items). Proximity card 1224 also includes a volume control affordance 1224C for controlling the volume of the Kitchen Speaker. In this example, the media content displayed in proximity card 1224 is content that is currently playing on the device 1200 (e.g., affordance 1224A), and other media items (e.g., affordance 1224D), for example, media that was recently played on the first device 1200. Proximity card 1224 includes identifier 1224B below affordance 1224A, which indicates which device (e.g., if any) is currently playing back the media item corresponding to affordance 1224A. In this example, identifier 1224B includes a graphic (e.g., which can be animated to convey that music is currently playing) and the text "From iPhone", which refers to device 1200 in this example. Proximity card 1224 provides an interface that allows the user to add the device "Kitchen Speaker" into the current playback session (e.g., of "Monday Morning" on device 1200).

[0442] As shown in FIG. 12I, proximity card 1224 also includes device selection affordance 1224E. In some examples, device selection affordance 1224E can be used to change the focus of the proximity card. For example, if the user desired to add the iPhone (e.g., device 1200) to the media session of the Kitchen Speaker (e.g., playback of the song "Same Thing" as shown in FIG. 12H), they can select affordance 1224E and change the targeted device (e.g., of proximity card 1224) to iPhone, and then select the indicator associated with the media item "Same Thing". For example, after receiving user input on 1224E and selection of the device "iPhone" (e.g., via a drop down menu, or the like), device 1200 updates the display of affordance 1224A to show the

album art of the song "Same Thing", and updates identifier 1224B to include the text "From Kitchen Speaker".

[0443] In some embodiments, a proximity card is displayed concurrently with a current media interface. In some embodiments, display of a proximity card replaces display of a portion of a current media interface. For example, FIG. 12I shows proximity card 1224 having replaced the album art 1220B of current media interface 1220 (of FIG. 12H). In some embodiments, a proximity card is not displayed concurrently with a current media interface. For example proximity card 1208 of FIG. 12E is not currently displayed with a current media interface (e.g., 1220 of FIG. 12H).

[0444] At FIG. 12J, device 1200 receives user input 1226 representing selection of affordance 1224A. In response to user input 1226, device 1200 displays current media interface 1228, as shown in FIG. 12K. Current media interface 1228 represents an updated version of interface 12H. As can been, device selection indicator 1228D has been updated to indicate that the Kitchen Speaker has been added to the media session for the song "Same Thing"—the accompanying text now states "iPhone + Kitchen". Media indicator 1228A identifies the song titled "Monday Morning" currently playing on device 1200 and device 1202, playback control affordances 1228B, album art 1228C, device selection affordance 1228D (indicating that device 1200 is targeted on itself, the device named "iPhone", and also targeted on the Kitchen device set).

[0445] FIGS. 12L-12N illustrate exemplary interfaces for controlling media playback on a second device. In some examples, affordance 1224E of FIG. 12I can be used to change the function of the proximity card. For instance, a user may desire to control playback of the existing media session that includes the second device (e.g., device 1202) rather than to add the second device to a media session of the first device (e.g., device 1200). For example, referring back to FIG. 12H, the user may want to maintain the two separate media sessions that are currently playing back on the two devices, but wants to control playback on the second device 1202. In some examples, the device receives selection of affordance 1224E of FIG. 12I and, in accordance with the selection (e.g., with our without additional user input), subsequently changes the displayed proximity card into a current media interface proximity card 1230, as shown in

FIG. 12L. In this example, proximity card 1230 of FIG. 12L replaces display of proximity card 1224 of FIG. 12I. Proximity card 1230 is a type of current media interface, but is presented visually as a proximity card (e.g., with a displayed border, and occupying less than all of the display 1206), to visually indicate that it represents a current media interface for a device that was placed in close proximity to device 1200. In some embodiments, in accordance with a selection of 1224E, device 1200 displays a current media interface such as that shown in FIG. 12H (e.g., full size).

[0446] Proximity card 1230 includes an identifier 1230A that indicates the current focus of the proximity card (e.g., Kitchen Speaker), a media identifier 1230B identifying a media item currently being played back by the targeted device (e.g., the song "Same Thing" by "Jambug"), and playback control affordances 1230C for controlling playback of the current media session.

[0447] At FIG. 12M, device 1200 receives user input 1232 associated with a playback control affordance 1230C. In this example, the user input is a selection "next track" affordance. User input 1232 can be a tap, a touch, a deep press, a tap and hold input, or the like.

[0448] At FIG. 12N, in response to user input 1232, device 1200 updates the display of media identifier 1230B to identify the next track (e.g., the song "Tuesday Morning" by "The Fomoers").

[0449] FIGS. 12O-12W illustrate an exemplary scenario in which only a first device is playing back media content when placed in close proximity with to a second device.

[0450] FIG. 12O illustrates an exemplary scenario in which a user is holding device 1200 (e.g., a personal device) in close proximity to device 1202 (e.g., a smart speaker). Thus, FIG. 12O is similar to FIG. 12A, the description of which is hereby incorporated by reference. In the example depicted in FIG. 12O, device 1200 is currently outputting audio, as represented by sound waves 1234 (from device 1200), and device 1202 is not currently outputting audio (e.g., no media playback).

[0451] FIG. 12P illustrates the state of device 1200 immediately prior to being placed in close proximity with device 1202. As can be seen, device 1200 is currently playing back a first media item, as indicated by current media interface 1220 (e.g., described above with respect to

FIG. 12H). Media indicator 1220A identifies the song titled "Monday Morning" currently playing on device 1200.

[0452] FIG. 12Q illustrates an exemplary media information interface. FIG. 12Q illustrates media information interface 1236 displayed on a display 1206 of device 1200. In some examples, device 1200 displays media information interface 1236 in accordance with detecting that device 1200 is in close proximity to device 1202. For example, media information interface 1236 is displayed in response to detecting that device 1200 is within (e.g., at least a portion of the device) the proximity zone 1204 (also referred to as proximity condition range indicator 1204) of device 1202. Thus, the proximity condition is satisfied.

[0453] In some examples, media information interface 1236 is displayed in response to device 1200 being placed (e.g., briefly or continuously) in close proximity with device 1202. Thus, because media information interface 1236 displayed in response to the device being placed in close proximity, it is also referred to as a "proximity card" or specifically a "media proximity card".

[0454] In the example shown in FIG. 12Q, device 1202 is a device associated with the location (or name) "Kitchen" and is a speaker (e.g., a smart speaker). Thus, proximity card 1236 includes the title 1236A "Kitchen Speaker", visually indicating the name of the device that the user placed device 1200 in close proximity to (e.g., also referred to as the "focus" or "target" of the proximity card). Proximity card 1236 also includes an action indicator 1236B (e.g., an arrow), indicating that a user input gesture on the proximity card (e.g., in the direction of the arrow) will cause the device to perform an action. Likewise, action indicator 1236D (the text "Push to Play") indicates to a user that a user input on the proximity card 1236 will cause the device to take action. Additionally, proximity card 1236 includes a media affordance 1236C representing media content (e.g., one or more media items). In this example, media affordance 1236C includes album art for the track titled "Monday Morning" by the artist "The Fomoers" (e.g., as shown in FIG. 12P). In some examples, media affordance 1236C includes other types of visual indications of one or more media items (e.g., text, images, or other appropriate visual representations). In this example, the media content represented (e.g., by affordance 1236C) by the proximity card 1236 is content that is currently playing on the device 1202 (e.g., the Kitchen

Speaker). Proximity card 1236 provides an interface that allows the user to add the device "Kitchen Speaker" to the current media session (e.g., of playback of "Monday Morning" on device 1200).

[0455] FIGS. 12R-12U illustrate and exemplary push interaction with a media control interface. At FIG. 12R, device 1200 receives user input 1238 on proximity card 1236. At FIG. 12S, user input 1238 has moved across the display (e.g., in the direction of action indicator 1236B). In this example, user input 1238 is a directional swipe gesture. FIG. 12R also illustrates proximity card 1236 has shifted position on the display 1206 in accordance with the directional swipe gesture user input 1238, which will be referred to as the "push position".

[0456] In some examples, device 1200 requires that the proximity card 1236 be held at the push position for a threshold amount of time before performing an action in response. In some examples, device 1200 provides a visual indication of the progress toward satisfying the threshold amount of time. For example, FIG. 12T illustrates push progress indicator 1236E, which is a partially completed ring. In some examples, other types of indicators can be used. Thus, in the example in FIG. 12T, the elapsed time that the proximity card 1236 has been held in the push position is approximately one-quarter of the necessary threshold time. In some embodiments, a requirement to satisfy a threshold amount of time in the push position can be bypassed (e.g., overridden) if a user input (e.g., 1238) satisfies an additional criteria. For example, if the user input is a deep press or is a swipe past an edge of the display (e.g., through the push position and off the top edge of display 1206), the device can perform the push function immediately in response.

[0457] In response to the user input 1238, device 1200 performs an action in response. In some embodiments, the device 1200 "pushes" media playback onto the device 1202. In some examples, device 1200 initiates playback of the media item represented by proximity card 1236 on device 1202 (e.g., transmits an instruction to device 1202 in response to user input 1238). In some examples, playback on device 1200 ceases in response to user input 1238. FIG. 12U illustrates proximity card 1236 upon completion of the push action (e.g., user input 1238 holding the proximity card 1236 in the push position for longer than the threshold duration). As shown, proximity card 1236 now includes confirmation 1236F ("Playing on Kitchen Speaker") and push

progress indicator 1236E now includes a checkmark. Thus, in this example, when the display 1206 of device 1200 appears as shown in FIG. 12U, the device 1202 is currently playing back the media item "Monday Morning".

[0458] FIG. 12 V illustrates an exemplary current media interface 1220. Current media interface 1220 is displayed subsequent to device 1200 receiving user input 1238 pushing the playback to the Kitchen Speaker. Thus, device 1200 updates the device selection indicator 1220C to indicate that the current focus of the device (e.g., the devices in the current media session) is the Kitchen Speaker. Thus, in this example, playback on the device iPhone (e.g., device 1200) ceased and audio playback was handed off to the Kitchen Speaker. In some embodiments, pushing playback includes initiating playback on the target device (e.g., hand off playback) and ceasing playback on the current device (e.g., 1200). In some embodiments, pushing playing includes initiating playback on the target device (e.g., hand off playback) and continuing playback on the current device (e.g., 1200).

[0459] FIG. 12W illustrates an alternate technique for pushing media to a second device, in accordance with some embodiments. FIG. 12W depicts current media interface 1220, as described with respect to FIG. 12P. Notably, device 1200 is currently targeted at itself ("iPhone"). Device 1200 receives user input 1239 on device selection affordance 1220C. In some examples, a multi-device interface is displayed in response to user input 1239. As described above, the multi-device interface can be used to add the second device (e.g., 1202) to the current media session (e.g., effectively pushing playback to it).

[0460] In accordance with some embodiments, media content is automatically transferred from a first device to a second device if the devices are in close proximity for a length of time greater than a threshold length of time. For example, if device 1200 is placed in close proximity to device 1202 for longer than a threshold length of time (e.g., 5 seconds), device 1200 can automatically (e.g., without further user input) transfer playback of media to device 1202. This can be referred to as "automatic push" or "auto-push". In accordance with some embodiments, media content is automatically transferred from a second device to a first device if the devices are in close proximity for a length of time greater than a threshold length of time. For example, if device 1200 is placed in close proximity to device 1202 for longer than a threshold length of

time (e.g., 5 seconds), device 1202 can automatically (e.g., without further user input) transfer playback of media to device 1200. This can be referred to as "automatic pull" or "auto-pull". When a user places a device in close proximity to another device for an extended period of time, this can be a strong indication that the user wishes playback to transfer between the devices. For example, a user can place their smartphone next to their smart speaker (e.g., 580) after entering their home and have what they were listening to on headphones begin playing on their home sound system, without further input at any device. Thus, by automatically pushing or pulling content from a personal electronic device, the number of user inputs for transferring media playback is reduced.

[0461] FIGS. 12X-12AG illustrate an exemplary scenario in which only a second device is playing back media content when placed in close proximity with to a first device.

[0462] FIG. 12X illustrates an exemplary scenario in which a user is holding device 1200 (e.g., a personal device) in close proximity to device 1202 (e.g., a smart speaker). Thus, FIG. 12X is similar to FIG. 12A, the description of which is hereby incorporated by reference. In the example depicted in FIG. 12X, device 1200 is not currently outputting audio, and device 1202 is currently outputting audio (e.g., media playback), as represented by sound waves 1240 (from device 1202).

[0463] FIG. 12Y illustrates an exemplary media information interface. In some examples, device 1200 displays media information interface 1242 in accordance with detecting that device 1200 is in close proximity to device 1202. For example, media information interface 1242 is displayed in response to detecting that device 1200 is within (e.g., at least a portion of the device is inside) the proximity zone 1204 (also referred to as proximity condition range indicator 1204) of device 1202. Thus, the proximity condition is satisfied.

[0464] FIG. 12Y illustrates media information interface 1242 displayed on a display 1206 of device 1200. In some examples, media information interface 1242 is displayed in response to device 1200 being placed (e.g., briefly or continuously) in close proximity with device 1202. Thus, because media information interface 1242 displayed in response to the device being placed in close proximity, it is also referred to as a "proximity card" or specifically a "media proximity card".

[0465] In the example shown in FIG. 12Y, proximity card 1242 includes the title 1242A "iPhone", visually indicating the name of the device (e.g., device 1200) that is the target of the proximity card action (e.g., will receive media playback in response to appropriate user input on the proximity card). Proximity card 1242 also includes an action indicator 1242C (e.g., an arrow), indicating that a user input gesture on the proximity card (e.g., in the direction of the arrow) will cause the device to perform an action. Likewise, action indicator 1242D (the text "Pull down to play on iPhone") indicates to a user that a user input on the proximity card 1242 will cause the device to take action. Additionally, proximity card 1242 includes a media affordance 1242B representing media content (e.g., one or more media items). In this example, media affordance 1242B includes album art for the track titled "Same Thing" by the artist "Jambug" (e.g., that is currently playing on device 1202). In some examples, media affordance 1242B includes other types of visual indications of one or more media items (e.g., text, images, or other appropriate visual representations). In this example, the media content represented (e.g., by affordance 1242B) by the proximity card 1242 is content that is currently playing on the device 1202 (e.g., the Kitchen Speaker). Proximity card 1242 provides an interface that allows the user to add the device "iPhone" (e.g., 1200) to the current media session (e.g., of playback of "Same Thing" on device 1202) of another device in close proximity (e.g., device 1202).

[0466] FIGS. 12Z-12AC illustrate and exemplary pull interaction with a media control interface. At FIG. 12Z, device 1200 receives user input 1244 on proximity card 1242. At FIG. 12AA, user input 1244 has moved across the display (e.g., in the direction of action indicator 1242C). In this example, display 1206 is a touch-sensitive display. In this example, user input 1244 is a directional swipe gesture. FIG. 12AA also illustrates proximity card 1242 has shifted position on the display 1206 in accordance with the directional swipe gesture user input 1244, which will be referred to as the "pull position".

[0467] In some examples, device 1200 requires that the proximity card 1242 be held at the pull position for a threshold amount of time before performing an action in response. In some examples, device 1200 provides a visual indication of the progress toward satisfying the threshold amount of time. For example, at FIG. 12AB illustrates pull progress indicator 1242E, which is a partially completed ring. In some examples, other types of indicators can be used. Thus, in the example in FIG. 12AB, the elapsed time that the proximity card 1242 has been held

in the push position is approximately one-quarter of the necessary threshold time. In some embodiments, a requirement to satisfy a threshold amount of time in the push position can be bypassed (e.g., overridden) if a user input (e.g., 1244) satisfies an additional criteria. For example, if the user input is a deep press or is a swipe past an edge of the display (e.g., through the push position and off the top edge of display 1206), the device can perform the pull function immediately in response.

[0468] In response to the user input 1242, device 1200 performs an action in response. In some embodiments, the device 1200 "pulls" media playback onto the device 1200. In some embodiments, pulling playback includes initiating playback on the current device (e.g., 1200) and ceasing playback on the external device (e.g., 1202). In some embodiments, pulling playback includes initiating playback on the current device (e.g., 1200) and continuing playback on the external device (e.g., 1202).

[0469] In some examples, device 1200 initiates playback of the media item represented by proximity card 1242. In some examples, playback on device 1202 ceases in response to user input 1242 or in response to device 1200 initiating playback. FIG. 12AC illustrates proximity card 1242 upon completion of the pull action. As shown, proximity card 1242 now includes confirmation 1242F ("Playing on iPhone") and pull progress indicator 1242E now includes a checkmark. Thus, in this example, when the display 1206 of device 1200 appears as shown in FIG. 12AC, the device 1200 is currently playing back the media item titled "Same Thing".

[0470] FIG. 12AD illustrates an exemplary current media interface 1246. Current media interface 1246 is displayed subsequent to device 1200 receiving user input 1244 and pulling the playback to the iPhone. Current media interface 1246 includes a media identifier 1246A (e.g., showing the track titled "Same Thing" by "Jambug"), device selection affordance 1246B (e.g., showing focus is on the iPhone), media playback indicator 1246C (showing temporal progress of playback of the current media item), and playback control affordances 1246D. As shown in FIG. 12AD, device 1200 updates the device selection indicator 1246B to indicate that the current focus of the device (e.g., the devices in the current media session) is the iPhone (e.g., device 1200). Thus, in this example playback on the device Kitchen Speaker (e.g., device 1202) ceased and audio playback was handed off to the iPhone (e.g., device 1200).

[0471] In some examples, if device 1200 is associated with a personal listening device (e.g., headphones), device 1200 automatically pulls the media playback from device 1202 and initiates the playback on the personal listening device. For example, as shown in the current media interface 1246 of FIG. 12AE, the playback of the pulled media item "Same Thing" is currently playing back on the headphones named "AirPods" (e.g., indicated by device selection indicator 1246B). For example, if a user is about to leave their home and would like to take music with them (e.g., to go running), device 1200 routes the music to an associated set of headphones (e.g., if turned on and/or being worn by the user) without first beginning playback on a speaker of device 1200 in response to user input 1244.

[0472] FIG. 12AF-12AG illustrates an exemplary alternate technique for interacting with a proximity card, in accordance with some embodiments. FIG. 12AF depicts proximity card 1242, with a user input 1248 on the media affordance 1242B. In some examples, user input 1248 is a tap, a touch, a deep press, or a press and hold input, or the like. At FIG. 12AG, device 1200 displays a current media interface 1250 that is targeted at the Kitchen Speaker in response to user input 1248. Current media interface 1250 includes a device selection affordance 1250A (e.g., indicating the focus of the interface 1250) and playback control affordances 1250B. Thus, if a user wishes to control playback on the Kitchen speaker, rather than pull the media to the iPhone, a tap on the proximity card can be used to reveal a current media interface targeted to the Kitchen Speaker's media session.

[0473] FIGS. 12AH-12AJ illustrate exemplary interfaces for transferring playback of video content between devices. In the examples described above, reference was made to transferring playback of media, and in particular audio. In some embodiments, media content includes video content. For example, a media item can be a video file or a video stream. Thus, the techniques described above are applicable to video content as well.

[0474] For example, FIG. 12AH illustrates playback of exemplary video content 1252 on device 1200. FIG. 12AI illustrates an exemplary proximity card 1254 for transferring video content from device 1200 to another device. Proximity card 1254 is analogous to proximity card 1236 of FIG. 12Q, the description of which is hereby incorporated by reference. Similar to proximity card 1236, proximity card 1254 includes the title 1254A "Apple TV", visually

indicating the name of the device that the user placed device 1200 in close proximity to. Title 1254A also visually indicates the focus of the proximity card (e.g., the device that will receive the media content in response), in this case the device named "Apple TV". Proximity card 1254 also includes an action indicator 1254D (e.g., an arrow), indicating that a user input gesture on the proximity card (e.g., in the direction of the arrow) will cause the device to perform an action. Likewise, action indicator 1254C (the text "Push to Play") indicates to a user that a user input on the proximity card 1254 will cause the device to take action. Additionally, proximity card 1254 includes a media affordance 1254B representing media content (e.g., one or more media items). In this example, media affordance 1254B includes a video screen shot (e.g., depicting a dog) of a video (e.g., 1252 of FIG. AH) currently playing on device 1200. In some examples, media affordance 1254B includes other types of visual indications of one or more media items (e.g., a video screen shot, video cover art, text, images, or other appropriate visual representations). Proximity card 1254 provides an interface that allows the user to add the device "Apple TV" to the current media session (e.g., of playback of the video media 1252).

[0475] FIG. 12AJ illustrates an exemplary result of transferring the video content depicted in FIG. 12AI from a first device 1200 to a second device 1260. In some examples, device 1260 can be a device as described above with respect to device 660 of FIG. 6. In this example, device 1260 is media device (e.g., Apple TV made by Apple Inc. of Cupertino, California USA) connected to a display 1256 of device 1258 (e.g., a television). In this example, the user placed device 1200 in close proximity to device 1260, causing device 1200 to display proximity card 1254. In response to user input representing an upward swipe on proximity card 1254, playback of the video content is transferred to the device 1260. For example, device 1260 is currently playing back the video content 1252 from FIG. 12AH. In this example, device 1260 is configured to output the video content via the display of television device 1256. In some embodiments, device 1260 includes a display, which is used to playback the transferred media content.

[0476] Further, in the example depicted in FIG. 12AJ, device 1202 is currently playing back the audio portion of the video content playing back on device 1260. In some examples, device 1260 acts like a media hub, and can route video and/or audio data (e.g., for playback and output) to other devices connected to it (e.g., via wireless or wired connections). For example, as shown

in FIG. 12AJ, device 1260 outputs video playback to television 1256 and audio playback (e.g., from the same media) to smart speaker 1202. In some embodiments, the device routing of a component signal (e.g., video or audio) of media content is preconfigured. For example, a user can configure device 1260 to always output audio on device 1202 and always output video on device 1256 (e.g., when a request for playback is received at either device, or a some other external device via a multi-device interface). In some embodiments, device 1260 dynamically determines the routing of a component signal (e.g., video or audio) of media content. For example, a user can configure device 1260 to always output audio on device 1202 and always output video on device 1256—however, if device 1202 is busy playing back other content, device 1260 can instead route the audio signal to be output by the speakers of television 1256. In some embodiments, device 1260 automatically detects the appropriate device (e.g., a preferred device) and automatically routes the content to that device. For example, device 1260 automatically detects speaker 1202 and automatically routes (e.g., hands off) audio playback of the audio portion of media). Likewise, device 1202, in some examples, automatically routes video to a video playback device, such as device 1260 (e.g., an Apple TV).

[0477] One of skill in the art would appreciate the numerous possible combinations of devices (e.g., two or more) that can be used to create an integrated media playback experience as described herein (e.g., with respect to FIGS. 12A-12AM). For example, the functions as described above (e.g., with respect to detecting proximity, displaying interfaces for transferring media between devices, displaying visual content, and/or outputting audible content) are not limited to being performed by a single device. In some embodiments, any number of devices are used to perform a function (e.g., and achieve a desired end result). One of skill in the art would appreciate that any number of devices can be used to perform the functions described above and, thus, to achieve the same desired end result, and the various combinations of devices possible are intended to be within the scope of this disclosure. In some examples, reducing the dependence on a single device or a limited number of devices to perform a desired media function, provides the user with a seamless and integrated media experience that can be accessed using various devices.

[0478] FIGS. 12AK-12AL illustrate exemplary interfaces for transferring a phone call between devices. In the examples described above, reference was made to transferring playback

of media, and in particular audio and video. However, for example, the transfer of playback is not limited to transferring media such as songs and movies. In some embodiments, media content includes a phone call. For example, media content can be a telephone call, a voice over internet protocol "VOIP" call, a video call, or the like.

[0479] For example, FIG. 12AK illustrates an exemplary call interface 1262, displayed on device 1200. Call interface 1262 indicates that device 1200 is currently in an active phone call (e.g., a VOIP call).

[0480] FIG. 12AL illustrates an exemplary proximity card 1264 for transferring the phone call to another device (e.g., 1202). Proximity card 1264 is analogous to proximity card 1236 of FIG. 12Q, the description of which is hereby incorporated. Similar to proximity card 1236, proximity card 1264 includes the title 1264A "Kitchen Speaker", visually indicating the name of the device that the user placed device 1200 in close proximity to. Title 1264A also visually indicates the focus of the proximity card (e.g., the device that will receive the media content in response), in this case the device named "Kitchen Speaker". Proximity card 1264 also includes an action indicator 1264C (e.g., an arrow), indicating that a user input gesture on the proximity card (e.g., in the direction of the arrow) will cause the device to perform an action. Likewise, action indicator 1264D (the text "Push Call") indicates to a user that a user input on the proximity card 1264 will cause the device to take action. Additionally, proximity card 1264 includes a media affordance 1264B representing media content (e.g., a phone call). In this example, media affordance 1264B includes an image of a contact, a contact name, and a call time for the current phone call. In some examples, media affordance 1264B includes any appropriate visual indication of one or more media items (e.g., a video screen shot, video cover art, text, images, or other appropriate visual representations). Proximity card 1264 provides an interface that allows the user to add the device "Kitchen Speaker" to the current media session (e.g., of outputting audio of the phone call).

[0481] As one of skill in the art would appreciate, a phone call involves both audio output and audio input (e.g., using a microphone). In some embodiments, transferring a media item (e.g., a phone call or a video call) from one device to another device comprises transferring media output (e.g., audio or video output), but not transferring a media input function (e.g., audio

or video input). For example, in the example depicted in FIG. 12AL, device 1200 can handoff audio output (e.g., so that phone call audio output is audibly output by a smart speaker 1202), but forgo handing off the media input function and continue to use its microphone to receive phone call audio (e.g., or its camera, in the case of a video call). In some embodiments, transferring a media item (e.g., a phone call or a video call) from one device to another device comprises transferring media output (e.g., audio or video output), and transferring a media input function. For example, in the example depicted in FIG. 12AL, device 1200 can handoff audio output (e.g., so that phone call audio output is audibly output by a smart speaker 1202), as well has handoff the media input function (e.g., so that smart speaker 1202 uses its microphone to receive phone call audio). In some embodiments, transferring a media item (e.g., a phone call or a video call) from one device to another device comprises transferring media input function, but not transferring a media output (e.g., audio or video output). In some embodiments, a device transfers less than all media output functions and/or media input functions. For example, if a video call is pushed to a display device with no audio output capability, the device (e.g., 1200) can transfer the visual output function to the display the device, and retain both the audio output function and the audio input function.

[0482] It should be apparent to one of skill in the art that the various interfaces and techniques described herein (e.g., with respect to FIGS. 12A-12AM) for controlling and/or transferring media content between devices is applicable to many types of media. Any media with one or more of an audio component and a visual component, and that are formatted to be communicated between electronic devices, are intended to be within the scope of this disclosure. For example, other specific examples of media that can be controlled or transferred in accordance with the techniques described herein include, but are not limited to: an electronic book (e.g., also commonly referred to as an "e-book"), a webpage, textual content (e.g., documents, spreadsheets), screen sharing (e.g., mirroring a first device's screen at a second device), and an image (e.g., photos, animations).

[0483] In certain situations, the owner of a device may not desire to allow any user that has physical access to a device to be able to control the device. For example, it may not be desirable to allow guests (e.g., via the guests' devices) to control playback on the owner's home device. FIG. 12AM illustrates an exemplary media information interface. Media information interface

1266 (also referred to as proximity card 1266) depicts an interface that is displayed, in some examples, when a guest device is placed in close proximity to a home owner's device. In some examples, a guest device is a device that is not associated with a user account that is a member of an authorized set of users (e.g., one or more). For example, an authorized set of users (e.g., user accounts) are users that have been added to a group in a device management application (e.g., a group of users 602A-602C as described above with respect to FIG. 6A). Thus, in some examples, if device 1200 is not logged into a user account associated with a user of the authorized set of users, device 1200 does not present a proximity card for controlling playback (e.g., a push proximity card, a pull proximity card, or a recently played media proximity card). Rather, in some examples, device 1200 displays proximity card 1266 in response to being placed in close proximity to a device that is associated with an authorized set of users.

[0484] Proximity card 1266 includes the title 1266A "Kitchen Speaker", visually indicating the name of the device (e.g., device 1202) that is the target of the proximity card (e.g., whose media playback information is presented on the card). Additionally, proximity card 1266 includes a media affordance 1266B representing media content (e.g., one or more media items). In this example, media affordance 1266B includes album art for the track titled "How Many Times" by the artist "Jambug" (e.g., that is currently playing on device 1202). In some examples, media affordance 1266B includes other types of visual indications of one or more media items (e.g., text, images, or other appropriate visual representations). In this example, the media content represented (e.g., by affordance 1266B) by the proximity card 1266 is content that is currently playing on the device 1202 (e.g., the Kitchen Speaker). Proximity card 1266 also includes a textual identification 1266C of the currently playing media. Thus, proximity card 1266 provides an informational interface that allows a guest to view information about media playback on a device in close proximity, without being able to affect the playback. For example, a swipe on proximity card 1266 will not push or pull media playback to device 1200.

[0485] FIG. 13 is a flow diagram illustrating a method for displaying a media information interface using an electronic device in accordance with some embodiments. Method 1300 is performed at a device (e.g., 100, 300, 500) with a display. Some operations in method 1300 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0486] As described below, method 1300 provides an intuitive way for displaying a media information interface. The method reduces the cognitive burden on a user for displaying a media information interface, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to display a media information interface faster and more efficiently conserves power and increases the time between battery charges.

[0487] The electronic device (e.g., 1200) detects (1302), while connected to a second device (e.g., 1202), an indication that a physical proximity between the first device and the second device satisfies a proximity condition. For example, the devices are connected when the second device is in Bluetooth range of the first device, or is connected to the same Wi-Fi network as the first device. In some examples, the proximity condition is satisfied when a physical proximity of the electronic device (e.g., 1200) is within a proximity zone of the second device (e.g., such as proximity zone 1204 of FIG. 12B). In some embodiments, the electronic device (e.g., device 1200 or 1202) detects whether the proximity condition is satisfied using a near-field communication technique. In some embodiments, the electronic device (e.g., 1200) that determines whether the proximity condition is satisfied communicates the determination to the other device (e.g., 1202).

[0488] In accordance with detecting that the physical proximity satisfies the proximity condition (e.g., the electronic device is within the proximity zone 1204 of FIG. 12B), the electronic device (e.g., 1200) displays (1304), on the display (e.g., 1206), a media information interface (e.g., 1208 of FIG. 12B).

[0489] Using a detection of physical proximity for displaying a media information interface allows the user to efficiently invoke a device-specific user interface with fewer required user inputs. Performing an operation when a set of conditions has been met without requiring further user input enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0490] In some embodiments, the indication is detected while the first device (e.g., 1200) is not playing media content and while the second device (e.g., 1202) is not playing media content.

In some embodiments, displaying the media information interface (e.g., 1208) comprises displaying, on the display (e.g., 1206), a first affordance (e.g., 1208A) representing a first media item that was recently played on the second device. In some embodiments, the electronic device (e.g., 1200) receives a first user input (e.g., 1210) representing selection of the first affordance. In response to receiving the first user input (e.g., 1210), the electronic device (e.g., 1200) transmits an instruction to the second device (e.g., 1202) to initiate playback of the first media item.

[0491] In some embodiments, the electronic device (e.g., 1200) receives a second user input (e.g., 1214 of FIGS. 12E and 12F) representing a directional swipe gesture on the first affordance (e.g., 1208A). In response to receiving the second user input (e.g., 1214), and while continuing to display the first affordance (e.g., 1208A of FIG. 12F), displaying, on the display (e.g., 1206), a second affordance (e.g., 1208C of FIG. 12F) representing a second media item that was recently played on the second device, wherein the second affordance was not displayed prior to receiving the second user input (e.g., 1214 of FIG. 12F).

[0492] In some embodiments, the indication is detected while the first device (e.g., 1200) is playing a third media item (e.g., 1220A of FIG. 12H) and while the second device (e.g., 1202) is playing a fourth media item (e.g., 1222 of FIG. 12H) different than the third media item. In some embodiments, displaying the media information interface (e.g., 1224 of FIG. 12I) comprises displaying, on the display (e.g., 1206), a third affordance (e.g., 1224A of FIG. 12I) associated with the third media item (e.g., 1220A) currently playing on the first device (e.g., 1200). In some embodiments, the electronic device (e.g., 1200) receives a third user input (e.g., 1226 of FIG. 12J) representing selection of the third affordance (e.g., 1224A). In response to receiving the third user input (e.g., 1226), the electronic device (e.g., 1200) transmits an instruction to the second device (e.g., 1202) to initiate playback of the third media item (e.g., 1220A).

[0493] In some embodiments, further in response to receiving the third user input (e.g., 1226), the electronic device (e.g., 1200) ceases playing the third media item (e.g., 1220A) on the first device (e.g., 1200).

[0494] In some embodiments, further in response to receiving the third user input (e.g., 1226), the electronic device (e.g., 1200) displays, on the display, playback control affordances (e.g., 1228B of FIG. 12K) that control the playback of the third media item (e.g., 1228A of FIG. 12K, or 1230C of FIG. 12L) on the second device (e.g., 1202).

[0495] In some embodiments, the indication is detected while the first device (e.g., 1200) is playing a fifth media item (e.g., 1220A of FIG. 12P) and while the second device (e.g., 1202) is not playing media content. In some embodiments, displaying the media information interface (e.g., 1236 of FIG. 12Q) comprises displaying, on the display, a fourth affordance (e.g., 1236C of FIG. 12Q) representing the fifth media item (e.g., represented by 1236C of FIG. 12Q). In some embodiments, the electronic device (e.g., 1200) receives a fourth user input (e.g., 1238 of FIG. 12R and 12S). In some embodiments, the electronic device (e.g., 1200) determines whether the fourth user input (e.g., 1238) represents a directional swipe gesture on the fourth affordance (e.g., 1236C). In some embodiments, the swipe gesture is an upward swipe. In some embodiments, the swipe gesture is a swipe in a direction toward the second device (e.g., 1202). For example, the swipe gesture would depend on the orientation of the first device (e.g., 1200) relative to the second device (e.g., 1202). In response to receiving the fourth user input (e.g., 1238), and in accordance with a determination that the fourth user input represents a directional swipe gesture on the fourth affordance (e.g., 1236C): the electronic device (e.g., 1200) transmits an instruction to the second device (e.g., 1202) to initiate playback of the fifth media item (e.g., represented by affordance 1236C of FIG. 12Q, or by 1220A of FIG. 12C). In some embodiments, further in response to the fourth user input (e.g., 1236C), and in accordance with a determination that the fourth user input represents a directional swipe gesture on the fourth affordance (e.g., 1236C), the electronic device (e.g., 1200) ceases playback, on the first device (e.g., 1200), of the fifth media item. In some embodiments, in response to receiving the fourth user input (e.g., 1236C), and in accordance with a determination that the fourth user input does not represent a directional swipe gesture on the fourth affordance (e.g., 1236C): the electronic device (e.g., 1200) forgoes transmitting an instruction to the second device (e.g., 1202) to initiate playback of the fifth media item. In some embodiments, further in response to receiving the fourth user input (e.g., 1236C), and in accordance with a determination that the fourth user input does not represent a directional swipe gesture on the fourth affordance (e.g., 1236C), the

electronic device (e.g., 1200) continues playback, on the first device (e.g., 1200), of the fifth media item (e.g., represented by affordance 1236C of FIG. 12Q, or by 1220A of FIG. 12C).

[0496] In some embodiments, the first device (e.g., 1200) ceases playback of the fifth media item at a playback time (e.g., as illustrated by 1220D of FIG. 12V) of the fifth media item (e.g., represented by affordance 1236C of FIG. 12Q, or by 1220A of FIG. 12C). In some embodiments, transmitting the instruction to the second device (e.g., 1202) to initiate playback of the fifth media item comprises transmitting an instruction to the second device to begin playback at approximately the playback time of the fifth media item. For example, so that the playback is perceived as continuous by a listener. For example, the first device instructs the second device to begin playback at the same time that the first device ceases playback of the fifth media item, a certain amount of time after the time that the first device ceases playback, or within a range of time based on the time that the first device ceases playback.

[0497] In some embodiments, displaying the media information interface comprises displaying, on the display (e.g., 1206), a fifth affordance (e.g., 1220C of FIG. 12W). In some embodiments, the electronic device (e.g., 1200) determines whether the fourth user input (e.g., 1239) represents a selection of the fifth affordance (e.g., 1220C of FIG. 12W. In some embodiments, in response to receiving the fourth user input, and in accordance with a determination that the fourth user input represents a selection of the fifth affordance, displaying, on the display (e.g., 1206), a multi-device interface (e.g., 610 of FIG. 6J). In some embodiments, further in response to receiving the fourth user input, and in accordance with a determination that the fourth user input does not represent a selection of the fifth affordance, the electronic device (e.g., 1200) forgoes displaying the multi-device interface.

[0498] In some embodiments, the indication is detected while the first device (e.g., 1200) is not playing media content and while the second device (e.g., 1202) is playing a sixth media item (e.g., represented by 1242B of FIG. 12Y). In some embodiments, displaying the media information interface (e.g., 1242 of FIG. 12Y) comprises displaying, on the display (e.g., 1206), a sixth affordance (e.g., 1242B of FIG. 12Y)) representing the sixth media item. In some embodiments, the electronic device (e.g., 1200) receives a fifth user input (e.g., 1244 of FIGS. 12Z and 12AA). In some embodiments, the electronic device (e.g., 1200) determines whether

the fifth user input represents a directional swipe gesture on the fifth affordance (e.g., 1242B). In some embodiments, the swipe gesture is a downward swipe. In some embodiments, the swipe gesture is a swipe in a direction away from the second device (e.g., 1202). For example, the directional swipe would depend on the orientation of the first device relative to the second device. In some embodiments, in response to receiving the fifth user input (e.g., 1244), and in accordance with a determination that the fifth user input represents a directional swipe gesture on the sixth affordance: the electronic device (e.g., 1200) transmits an instruction to the second device (e.g., 1202) to cease playback of the sixth media item. In some embodiments, further in response to receiving the fifth user input (e.g., 1244), and in accordance with a determination that the fifth user input represents a directional swipe gesture on the sixth affordance, electronic device (e.g., 1200) initiates playback, on the first device (e.g., 1200), of the sixth media item (e.g., represented by 1242B of FIG. 12AA). For example, pulling on a displayed proximity card causes playback of media from the second device to transfer to the first device. In some embodiments, in response to receiving the fifth user input, and in accordance with a determination that the fifth user input does not represent a directional swipe gesture on the sixth affordance: the electronic device (e.g., 1200) forgoes transmitting an instruction to the second device (e.g., 1202) to cease playback of the sixth media item. In some embodiments, further in response to receiving the fifth user input, and in accordance with a determination that the fifth user input does not represent a directional swipe gesture on the sixth affordance: the electronic device (e.g., 1200) forgoes initiating playback, on the first device (e.g., 1200), of the sixth media item.

[0499] In some embodiments, the electronic device (e.g., 1200) determines whether the fifth user input (e.g., 1248 of FIG. 12AF) represents a selection (e.g., a tap, touch, press, deep press, or tap and hold input) of the sixth affordance (e.g., 1242B of FIG. 12AF). In some embodiments, further in response to receiving the fifth user input, and in accordance with a determination that the fifth user input (e.g., 1248) represents a selection of the sixth affordance, the electronic device (e.g., 1200) displays, on the display (e.g., 1206), playback control affordances (e.g., 1250B of FIG. 12AG) for controlling the playback of the sixth media item (e.g., represented by 1242B of FIG. 12AF) on the second device (e.g., 1202, represented by 1250A of FIG. 12AG). In some embodiments, in accordance with a determination that the fifth user input does not represent a selection of the sixth affordance, the electronic device (e.g., 1200)

forgoes displaying, on the display (e.g., 1206), playback control affordances (e.g., 1250B of FIG. 12AG) for controlling the playback of the sixth media item on the second device (e.g., 1202).

[0500] In some embodiments, the second device (e.g., 1202) ceases playback of the sixth media item (e.g., represented by 1246A of FIG. 12AD) at a playback time (e.g., 1246C of FIG. 12AD) of the sixth media item. In some embodiments, initiating playback, on the first device (e.g., 1200), of the sixth media item comprises beginning playback at approximately the playback time (e.g., 1246C of FIG. 12AD) of the sixth media item. In some examples, the transfer of playback from the second device to the first device is perceived as continuous by a listener. For example, the first device can begin playback at the same time that the second device ceases playback, a certain amount of time after the time that the second device ceases playback, or within a range of time based on the time that the second device ceases playback.

[0501] In some embodiments, initiating playback, on the first device (e.g., 1200), of the sixth media item (e.g., represented by 1246A of FIG. 12AD) comprises displaying, on the display (e.g., 1206): an identification of the sixth media item (e.g., 1246A of FIG. 12AD), and playback control affordances (e.g., 1246D of FIG. 12AD) for controlling playback of the sixth media item on the first device (e.g., 1200).

[0502] In some embodiments, the first device (e.g., 1200) is connected to a plurality of devices, wherein the plurality of devices includes the second device (e.g., 1202), and wherein the displayed media information interface (e.g., 1208, 1224, 1236, or 1242) corresponds to the second device (e.g., 1202).

[0503] In some embodiments, the first device (e.g., 1200) is connected to the second device (e.g., 1202) via a wireless communication link.

[0504] In some embodiments, the media information interface (e.g., 1236, or 1242) comprises an indication (e.g., 1236B, 1236D, 1242C, and/or1242D) that the media content playing back on the first device (e.g., 1200) or the second device (e.g., 1202) can be transferred to the other device not playing back the media content. For example, proximity cards can include an indication (e.g., text, an arrow) that the media content can be "pushed" to the second device (e.g., an up arrow) or "pulled" from the second device (e.g., down arrow).

[0505] In some embodiments, a media item is selected from the group consisting of an audio file, an audio stream, a video file, a video stream, a phone call, and a video call.

[0506] In some embodiments, the first device (e.g., 1200) is associated with a first user account (e.g., user 602A of FIG. 6A), and the second device (e.g., 1202) is associated with a set of user accounts. In some embodiments, the set of user accounts includes one or more user accounts. In some embodiments, further in accordance with detecting that the physical proximity satisfies the proximity condition: in accordance with a determination that the first user account is a member of the set of user accounts, the electronic device (e.g., 1200) displays an indication (e.g., 1236B, 1236D, 1242C, and/or1242D) that media playback can be transferred between the first device (e.g., 1200) and the second device (e.g., 1202). For example, the indication provides a visual indication that media playback can be pushed from the first device to the second, or pulled form the second device to the first. In some embodiments, the device also displays playback control affordances (e.g., 1220D of FIG. 12H). In some embodiments, in accordance with a determination that the first user account is not a member of the set of user accounts, the electronic device (e.g., 1200) forgoes displaying the indication (e.g., 1236B, 1236D, 1242C, and/or1242D) that media playback can be transferred between the first device and the second device. For example, the electronic device can instead display an interface 1266 as shown in FIG. 12AM. In some examples, interface 1266 includes an indication of what is now playing on the second device, but does not include an affordance that can be swiped to transfer playback, nor does it include arrows prompting a swipe gesture. In some embodiments, the device also forgoes displaying playback controls (e.g., as shown in FIG. 1266).

[0507] In some embodiments, in accordance with a determination (e.g., by device 1200 or 1202) that the physical proximity between the first device (e.g., 1200) and the second device (e.g., 1202) has satisfied the proximity condition continuously for a threshold amount of time, automatically transferring playback of a media item between the first device and the second device. In some embodiments, automatically transferring comprises transferring without further user input and/or without requiring an additional condition be met. In some embodiments, in accordance with a determination that the physical proximity between the first device and the second device has not satisfied the proximity condition continuously for a threshold amount of

time, the electronic device (e.g., 1200) forgoes automatically transferring playback of a media item between the first device and the second device.

[0508] Note that details of the processes described above with respect to method 1300 (e.g., FIG. 13) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 1300. For brevity, these details are not repeated below.

[0509] FIGS. 14A-14M illustrate exemplary techniques for controlling a device, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 15.

[0510] FIG. 14A illustrates an exemplary device 1400, in accordance with some embodiments. In some embodiments, device 1400 includes one or more features of devices 100, 300, 500, or 580. In some embodiments, as shown in FIG. 14A, device 1400 includes a touch-sensitive surface 1402. In this example, the touch-sensitive surface is arranged on top of the device, which is a smart speaker device sitting on surface. In some embodiments, device 1400 includes a touch-sensitive display.

[0511] FIG. 14B illustrates an overhead view of exemplary device 1400. Touch-sensitive surface 1402 can be seen. Also visible are LED indicators 1404, 1406A, and 1406B. In some examples, one or more of the LED indicators on device 1400 are used to visually indicate a status of the device. For example, touch-sensitive surface 1402 can respond differently to the same user input (e.g., a tap at the same location) based on a state of device 1400 (e.g., at different times). Thus, device 1400 can use the LED indicators to visually communicate the current status of the device (e.g., and in turn, indicate the effect that a user input (e.g., on touch-sensitive surface 1402) will have on the device). In some examples, one or more of the following properties of one or more LED indicators are controlled by the device 1400 to communicate a state of the device: illumination (e.g., on or off), brightness, color, diameter, and shape. In some examples, other visual properties of the LED indicators can be controlled (e.g., adjusted) in order to indicate a device state (e.g., a media playback status, a state of a touch-sensitive surface or a touch-sensitive display). In some embodiments, the LED indicators are a display screen (e.g.,

112 of FIG. 1B, 504 of FIG. 5A). For example, instead of using LED indicators, a device 1400 can include a display screen, such as an organic light-emitting diode (OLED) panel, an LED panel, an LCD panel, or the like. In some embodiments, LED indicators are any component(s) that can be controlled by an electronic device (e.g., 1400) to output a visual indication.

[0512] FIGS. 14C-14L depict various exemplary states of device 1400, and actions performed by the device in response to user input in accordance with the states.

[0513] FIG. 14C illustrates the overhead view shown in FIG. 14B, but where device 1400 receives a user input 1408 at a location on the touch-sensitive surface 1402. In this example, the user input 1408 is received at a location over LED indicator 1406A. In this example, device 1400 is not playing back media content at the time when user input 1408 is received. For example, device 1400 is in a paused state during a media playback session, or is not currently a part of a media playback session.

[0514] At FIG. 14D, in response to receiving user input 1408, device 1400 initiates playback of media. For example, device is now outputting media (e.g., via a speaker), as represented graphically by sound wave 1410 in FIG. 14D. In this example, device 1400 has resumed playing media playback that was paused. In some examples, if no media playback is currently paused, device 1400 still initiates playback of media in response to user input 1408. For example, the device 1400 can initiate playback of recently played media, default media (e.g., configured by user, or automatically selected), favorite media associated with a user account, preconfigured media (e.g., configured to playback during a predetermined time, day), or the like. In some examples, device 1400 updates the LED indicators to visually indicate that the device is currently playing back media content. In this example, indicator 1406A now includes a "+" icon and indicator 1406B now includes a "-" icon. In some examples, the indicators 1406A and 1406B of FIG. 14D visually communicate that the device is currently playing back media (e.g., outputting an audio signal). One of skill would appreciate that other visual indications (e.g., icons) and/or indicators (e.g., indicator 1406) can be used to visually communicate that the device is currently playing back media. Further, in some examples, indicators 1406A and 1406B of FIG. 14D visually communicate that they can be touched in order to adjust volume. One of skill would appreciate that other indications (e.g., icons) and/or indicators (e.g., indicator 1406)

can be used to visually communicate the state (e.g., function in response to user input) of a touch-sensitive or touch-sensitive display.

[0515] In some embodiments, the "+" and "-" indicators 1406A and 1406B cease to be displayed in response to device 1400 performing a voice assistant function. For example, in response to detecting voice user input (e.g., "Hey Siri"), the device can cease displaying one or more of the icons within indicator 1406A and 1406B. In some examples, the illumination of one or more of indicators 1404, 1406A, and 1406B changes during a voice assistant function. For example, indicators 1404, 1406A, and 1406B can turn off, be illuminated, change color, or the like in response to a voice assistant function being performed on device 1400.

[0516] FIG. 14E illustrates the overhead view shown in FIG. 14D, but where device 1400 receives a user input 1410 at a location on the touch-sensitive surface 1402. In this example, the user input 1410 is received at a location over LED indicator 1406A, which is the same location as user input 1408. In this example, device 1400 (as shown in FIG. 14D) is currently playing back media content at the time when user input 1410 is received. For example, device 1400 is included in a media playback session (e.g., and is playing music audio, or the audio component of video playback).

[0517] In response to receiving user input 1410, device 1400 adjusts a volume of device 1400. For example, as illustrated by the volume level meter 1412 in FIG. 14F, device 1400 adjusts (e.g., increases) volume from a first level 1412A (e.g., the volume immediately prior to the user input 1410) to a second level 1412B (e.g., the volume immediately subsequent to the user input 1410), in response to user input 1410.

[0518] FIGS. 14G-14J illustrate exemplary techniques for expanding the size of a responsive area on a touch-sensitive surface, in accordance with some embodiments.

[0519] FIG. 14G illustrates an overhead view of device 1400 while it is currently playing back media content (e.g., as described in FIG. 14F). Additionally, FIG. 14G depicts the boundaries of touch-sensitive regions 1414A and 1414B over indicators 1406A and 1406B, respectively. In some examples, touch-sensitive region boundaries are not displayed or visible on an electronic device (e.g., 1400). For instance, in this example, the regions 1414A and 1414B

are included for reference only and are not displayed by device 1400. In some examples, an electronic device displays boundaries of one or more touch-sensitive regions (e.g., 1414A or1414B).

[0520] At FIG. 14H, the device (as shown in FIG. 14G) receives user input 1416 over indicator 1406A. In this example, user input 1416 is a touch on touch-sensitive surface 1402. As can be seen, at least a portion of user input 1416 is included within the touch-sensitive region 1414A. In response to user input 1416, device 1400 increases a volume (e.g., as shown in FIG. 14F).

[0521] In some embodiments, further in response to receiving user input 1416 on an indicator 1406A, device 1400 temporarily expands the touch-sensitive region 1414A to form an expanded touch-sensitive region (also referred to as an "expanded region"). An exemplary expanded region 1418 is shown in FIG. 14I. As can be seen in FIG. 14I, the expanded region 1418 covers are larger area of touch-sensitive surface 1402 than region 1414A did (e.g., as shown in FIG. 14G). In some examples, an expanded region replaces a region. For instance, in this example, expanded region 1418 (e.g., which will perform the same function as region 1414A, in response to a touch within it) covers the area (and more) that region 1414A occupied, thus replacing it.

[0522] In some embodiments, the device ceases to maintain the expanded touch-sensitive region after a predetermined amount of time. In some embodiments, the device ceases to maintain the expanded touch-sensitive region after a predetermined amount of time since the last user input within the expanded region. For example, subsequent user inputs (e.g., taps) within the expanded region would cause a measure (e.g., a timer) of the amount of time since the last user input to be reset after each detect touch input. In this example, the expanded region returns to its normal size (e.g., 1414A) after the amount of time reaches the predetermined amount of time after the final user input with the expanded region.

[0523] At FIG. 14J, device 1400 receives a second user input 1420 within the expanded region 1418, but outside of region 1414A. In some examples, in accordance with a determination that the second user input (e.g., 1420) was received during the predetermined amount of time after detecting the first user input (e.g., 1416), device 1400 adjusts (e.g.,

increases) a volume of the currently playing media. For example, as illustrated by the volume level meter 1422 in FIG. 14K, device 1400 adjusts volume from a first level 1422A (e.g., the volume immediately prior to the user input 1420) to a second level 1422B (e.g., the volume immediately subsequent to the user input 1420), in response to user input 1420.

[0524] In some examples, in accordance with a determination that the second user input (e.g., 1420) was not received during the predetermined amount of time after detecting the first touch, device 1400 ceases playing media. For example, if user input 1420 was received after a predetermined amount of time after a touch that caused the region 1414A to expand (or a subsequent touch within the expanded region 1418), and was not within the region 1414A, then the volume is not increased because the temporary expansion (e.g., increase in responsive area on a touch-sensitive surface) ended before the user input 1420 was received. In some examples, device 1400 ceases playback in response to user input 1420, when it is received after the predetermined amount of time, and is outside of the second region. For instance, in the example shown in FIG. 14J, user input 1420 is within a region 1414C that is configured to cause device 1400 to pause media playback (or resume paused playback). Thus, in response to user input 1420 when it is received after the predetermined amount of time, and is outside of the second region, device 1400 pauses playback.

[0525] FIG. 14L illustrates an exemplary technique for adjusting the volume of a device. While the description above referenced increasing a volume of device 1400 in response to user input on a touch-sensitive surface, FIG. 19L illustrates an exemplary technique for lowering the volume of device 1400. In response to receiving user input 1410, device 1400 adjusts a volume of device 1400. For example, as illustrated by the volume level meter 1426 in FIG. 14L, device 1400 adjusts volume from a first level 1426A (e.g., the volume immediately prior to the user input 1424) to a second level 1426B (e.g., the volume immediately subsequent to the user input 1424, lower than level 1426A), in response to user input 1424.

[0526] In some embodiments, device 1400 includes a plurality of volume levels. For example, device 1400 can maintain e.g., store in memory) a volume level for a media playback function, and maintain a volume level for a voice assistant function. In some examples, the plurality of volume levels are each different. For example, device 1400 can adjust a first volume

level (e.g., of media playback) to 0% of maximum (e.g., muted, no sound is produced by a speaker of device 1400) while a second volume level is set to a non-zero value of 50% of maximum (e.g., a maximum output level of the speaker, or a maximum level of a volume level meter).

[0527] For example, FIG. 14M illustrates two volume level meters, representing two volume levels stored by device 1400. Media volume level meter 1430 includes a maximum level and a minimum level. Voice assistant volume level meter 1432 includes a maximum level and a minimum level.

[0528] In some embodiments, a second volume level is a derivative of a first volume level. For example, the volume level of a voice assistant (e.g., represented by volume meter 1432) can be mathematically derived from the volume level of media playback volume. For example, Volume (voice assistant) = (Volume (media) ÷ 1.5). Thus, the second volume level (e.g., of the voice assistant) is derived from a first volume level (e.g., of media playback), and adjustment of the first results in an adjustment of the second.

[0529] In some embodiments, a second volume level is non-zero when a first volume level is zero. For example, as shown in FIG. 14M, the virtual assistant volume meter 1432 has a non-zero minimum. Thus, in some examples, when media volume level is set to zero, the voice assistant volume is non-zero, and thus a user can continue to hear and interact with the voice assistant. For example, the voice assistant volume can be derived as follows (where Volume (voice assistant) = ((Volume (media) ÷ 1.5) + x), where "x" is a constant representing an amount of volume added back into the result of Volume(media) ÷ 1.5. One of skill in the art would appreciate that a volume level can be represented using a variety of units or quantities, which are not discussed in more detail here. Any appropriate representation of a volume level that an electronic device (e.g., device 1400) can use to control the volume of sound output is intended to be within the scope of this disclosure.

[0530] In some embodiments, user input on touch-sensitive surface 1402 causes device 1400 to change playback of currently playing media. For example, user input on touch-sensitive surface 1402 can cause device 1400 to skip to next media item (e.g., on a playlist), to go back to a previous media item (e.g., on a playlist), pause playback, or the like. In some embodiments,

device 1400 concurrently plays back a plurality of media items. In some embodiments, in response to user input on touch-sensitive-surface 1402, device 1400 adjusts the output of each of the plurality of concurrently playing media items. In some embodiments, each of the plurality of media items are adjusted differently. For example, if device 1400 is concurrently playing back first media representing a news program (e.g., an audio stream) and second media representing music (e.g., an audio stream), a user input (e.g., a swipe gesture, a rotational swipe, a multi-contact swipe, or the like) on surface 1402 causes device 1400 to adjust an output level of the first and second media. In some embodiments, the adjustment is relative. For example, an output level of the first media is adjusted higher while an output level of the second media is adjusted lower. For example, an output level of the first media is adjusted lower while an output level of the second media is adjusted higher. For example, if the user desires to increase the volume of the news program and lower the music volume, a counter-clockwise rotational swipe gesture on surface 1402 causes device 1400 to adjust the volume of each media item accordingly. Conversely, in this example, a clockwise rotational swipe gesture on surface 1402 causes device 1400 to increase the volume of the music and lower the volume of the news program. Thus, device 1400 receives user input to control a crossfade between a plurality of currently playing media signals.

[0531] In some embodiments, device 1400 includes one or more microphones. In some examples, device 1400 detects, with the one or more microphones, a voice command user input from a voice command source. In some embodiments, the voice command source is moving. For example, the voice command source is a user that is speaking to the device, and the user may move (e.g., walk around a room) relative to the device (e.g., stationary in the room). In some embodiments, while continuing to detect the voice command user input: device 1400 determines a plurality of spatial positions of the moving voice command source, and illuminates light-emitting elements (e.g., indicators 1404) based on the plurality of spatial positions of the moving voice command source. Thus, for example, device 1400 can use (e.g., selectively illuminate) indicators 1404 based on the direction of a voice command source (e.g., a user speaking to device 1400 to invoke a voice assistant function using device 1400). For instance, when a user stands at a first position relative to device 1400 (e.g., and speaks a command to the device), indicators 1404 include one or more illuminated LEDs that visually appear to align with or indicate (e.g., "point in") the direction of the first position. Likewise, when a user stands at a

second position relative to device 1400 (e.g., and speaks a command to the device), different than the first position, indicators 1404 include one or more illuminated LEDs that visually appear to align with or indicate (e.g., "point in") the direction of the second position.

[0532] In some embodiments, device 1400 detects a plurality of voice sources. In some embodiments, the device determines that one voice source, of the plurality of voice sources, represents a voice command user input. In some embodiments, device 1400 illuminates light-emitting elements (e.g., indicators 1404) based on the one voice source, of the plurality of voice sources. For example, if device 1400 is placed in a room with a plurality of speaking users and one user (or more than one) utters a voice command, the device 1400 uses indicators 1404 to visually indicate the direction of the user whose voice is currently being used to generate and/or process a voice command.

[0533] FIGS. 14N-14Q illustrate exemplary patterns created by illuminate light-emitting elements (e.g., indicators 1404) based on a spatial positions of a voice command source (e.g., 1440). At FIG. 14N, a user 1440 utters the user input voice command 1442 "Hey Siri" to invoke a personal assistant voice function on device 1400. FIG. 14O illustrates an overhead view (not to scale) of FIG. 14N, before the user utters voice command 1442. As shown, the indicator lights 1404 are not illuminated.

[0534] In FIG. 14P, an "input gravity" type of illumination pattern is shown, displayed in response to device 1400 detecting voice command 1442. The illuminated light-emitting elements of indicator 1404 form a semi-circular illumination pattern bordering the edge of indicators 1404. The elements that are closest to the edge and in the direction of the voice command source appear different from the surrounding elements. For example, the center of the semi-circle indicates the direction of the voice source (e.g., user 1440) being listened to by device 1400 and can appear to be brighter or a different color than the surrounding elements.

[0535] In FIG. 14Q, a "linear" type of illumination is shown (an alternative to FIG. 14P's "input gravity" type of illumination pattern), displayed in response to device 1400 detecting voice command 1442. The illuminated light-emitting elements of indicator 1404 form a substantially linear pattern within indicators 1404. Elements that align with the direction (e.g., relative to device 1400) of the voice command source appear different from the surrounding

elements. For example, the elements aligned with the direction of the voice source being listened to by device 1400 and can appear to be brighter or a different color than the surrounding elements.

[0536] In some examples, other appropriate types of visual patterns can be used to indicate direction of a voice command source, and the examples above are not intended to limit the scope of this disclosure.

[0537] FIGS. 14R-14T illustrate exemplary arrangements and states of light-emitting indicators, in accordance with some embodiments. For example, FIGS. 14R-14T illustrate exemplary illumination patterns of indicator 1404, for communicating various states of device 1400. In some embodiments, device 1400 includes a cover over indicator 1404. For example, a partially opaque cover placed over indicator 1404 can cause the output of the light emitting elements to diffuse, and form patterns, which are also illustrated in FIGS. 14R-14T, for ease of reference.

[0538] FIG. 14R illustrates four illumination states of indicator 1404, both with and without a diffusion cover over the indicator. Pattern 1454A illustrates a large pattern (all LEDs illuminated), pattern 1454B illustrates a medium pattern (middle two rings illuminated), pattern 1454C illustrates a small pattern (e.g., inner-most ring illuminated), and pattern 1454D illustrates a circling pattern (e.g., various LEDs illuminated in a shape). In some examples, lighting elements of 1404 form the shape of 1454D and alternate illumination to give the appearance that the shape is moving (e.g., swirling a circular pattern). Covered patterns 1452A-1452D correspond to patterns 1454A-1454D, respectively, and depict how the patterns appear with a diffusion cover over indicator 1404.

[0539] FIG. 14S illustrates four illumination patterns that are presented by indicator 1404 in response to a voice user input (e.g., invoking a voice assistant), in accordance with some embodiments. As shown in FIG. 14S, pattern 1454E is displayed prior to voice input. Patterns 1454F-1454I are displayed in sequence in response to the voice input, giving the appearance of a small pattern transitioning into a large pattern, and then settling at a medium sized pattern. Covered patterns 1452E-1452I correspond to patterns 1454E-1454I, respectively, and depict how the patterns appear with a diffusion cover over indicator 1404. In some embodiments, the

patterns 1454E-1454I form a series (e.g., that are displayed sequentially, in the order shown or in any order).

[0540] FIG. 14T illustrates three illumination patterns that are presented by indicator 1404, in accordance with some embodiments. For example, the patterns 1454J-1454L (and corresponding 1452J-1452L) can be displayed while device 1400 is performing a voice assistant function. In some embodiments, the patterns 1454J-1454L form a series (e.g., that are displayed sequentially, in the order shown or in any order). For example, performing a voice assistant function can include one or more of: listening to voice input (e.g., after detecting it), listening for voice input (e.g., prior to detecting it), outputting a voice assistant output (e.g., providing a text-to-speech audible response to voice input), processing user input (e.g., parsing the voice input, or performing a task in response to the voice input), or the like. In the examples depicted in FIG. 14T, each pattern includes shapes (e.g., circles) and each shape includes a number (e.g., the pattern in 1454J includes the numbers 1, 2, and, 3), wherein each number represents a color. Thus, the three circles of pattern 145J are each a different color. In patterns 1452J and 1454J, the elements in indicator 1404 are illuminated so that the colors do not appear to mix. In pattern 1452K and 1454K, the elements in indicator 1404 are illuminated so that the colors do appear to mix. For example, the colors are mixed and form new colors 4, 5, 6, and 7, based on the overlap of colors 1, 2, and 3. In pattern 1452L and 1454L, the elements in indicator 1404 are illuminated so that the colors do appear to mix, and the shapes within the pattern are less clearly defined (e.g., are no longer perfect circles, are swirling in an animated fashion). As can be seen, the colors 1 and 3 have changed relative position from pattern 1454K to 1454L. In some examples, the various patterns 1452J-1452L are be used (e.g., as sequence or part of an animation) to indicate the state of a voice assistant function.

[0541] In some examples, the color elements in indicator 1404, as described above with respect to patterns 1454J-1454L, are used to provide visual indication that the device 1400 is performing a first function (e.g., a voice assistant function), in contrast to a second function (e.g., media playback). For example, monochromatic light can be output by indicator 1404 when device 1400 is only playback media (e.g., a song), but transition to color when the device (e.g., even while continuing media playback) begins performing a voice assistant-related function (e.g., detects voice input of a token phrase, such as “Hey Siri”).

[0542] FIG. 15 is a flow diagram illustrating a method for controlling media playback using an electronic device in accordance with some embodiments. Method 1500 is performed at a device (e.g., 100, 300, 500) with a touch-sensitive surface. Some operations in method 1500 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0543] As described below, method 1500 provides an intuitive way for controlling media playback. The method reduces the cognitive burden on a user for controlling media playback, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to control media playback faster and more efficiently conserves power and increases the time between battery charges.

[0544] The electronic device (e.g., 1400) detects (1502) a touch (e.g., 1408 or 1410) on the touch-sensitive surface (e.g., 1402). In response to detecting the touch on the touch-sensitive surface, the electronic device (e.g., 1400) determines (1504) whether the electronic device (e.g., 1400) is currently playing media (e.g., represented by 1410 of FIG. 14D).

[0545] In accordance with a determination that the device (e.g., 1400) is currently playing media (1506): the electronic device (e.g., 1400) determines (1508) whether a location of the touch (e.g., 1408 or 1410) is within a first area (e.g., 1414C of the touch-sensitive surface (e.g.,1402) or within a second area (e.g., 1414A) of the touch-sensitive surface, wherein the first area and the second area are non-overlapping.

[0546] In accordance with a determination that the location of the touch is within the first area (e.g., 1414C) of the touch-sensitive surface, the electronic device (e.g., 1400) ceases playing media (1510).

[0547] In accordance with a determination that the location of the touch is within the second area (e.g., 1414A) of the touch-sensitive surface, the electronic device (e.g., 1400) adjusts (1512) a volume of the currently playing media (e.g., adjust up or down).

[0548] In accordance with a determination that the device is not currently playing media, the electronic device (e.g., 1400) initiates playback of media (1514).

[0549] Using a device state as well as a location of an area of a touch on a touch-sensitive surface to determine how a device responds to touch user input provides the user with a simplified interface that provides contextually relevant controls. Providing additional control of the device without cluttering the UI with additional displayed controls enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0550] In some embodiments, the second area is comprised of a plurality of non-contiguous and non-overlapping areas (e.g., 1414A and 1414B of FIG. 14G) of the touch-sensitive surface (e.g., 1402 of FIG. 14G).

[0551] In some embodiments, the second area is comprised of a first sub-area (e.g., 1414A) and a second sub-area (e.g., 1414B), and wherein adjusting the volume of the currently playing media comprises: lowering the volume of the currently playing media if the location of the touch is within the first sub-area (e.g., 1414B); and raising the volume of the currently playing media if the location of the touch is within the second sub-area (e.g., 1414A).

[0552] In some embodiments, the touch is a first touch (e.g., 1416), and, subsequent to adjusting the volume of the currently playing media in accordance a determination that the location of the touch (e.g., 1416) is within the second area (e.g., 1414A): the electronic device (e.g., 1400) temporarily expands the second area to create an expanded second area (e.g., 1418 of FIG. 14I) for a predetermined amount of time after detecting the first touch (e.g., 1416) on the touch-sensitive surface (e.g., 1402). In some embodiments, the electronic device (e.g., 1400) receives a second touch (e.g., 1420) on the touch-sensitive surface. In some embodiments, the second touch (e.g., 1420) is located within the expanded second area (e.g., 1418). In some embodiments, the second touch (e.g., 1420) is not located within the second area (e.g., 1414A). In some embodiments, the electronic device (e.g., 1400) determines whether the second touch (e.g., 1420) was received during the predetermined amount of time after detecting the first touch (e.g., 1416). In accordance with a determination that the second touch was received during the predetermined amount of time after detecting the first touch, the electronic device (e.g., 1400)

adjusts the volume of the currently playing media (e.g., as illustrated by 1422 of FIG. 14K). In accordance with a determination that the second touch was not received during the predetermined amount of time after detecting the first touch, the electronic device (e.g., 1400) ceases playing media. In some embodiments, ceasing playing media comprises pausing playback of currently playing media.

[0553] In some embodiments, the touch-sensitive surface includes a display (e.g., 1404A and 1404B). In some embodiments, the display is an LED array (e.g., 1406). In some embodiments, further in accordance with a determination that the device is currently playing media, the electronic device (e.g., 1400) displays, on the display, an indicator (e.g., + icon in 1406A of FIG. 14D) within the second area (e.g., 1414A). In some embodiments, further in accordance with a determination that the device is not currently playing media, the electronic device (e.g., 1400) forgoes displaying the indicator within the second area.

[0554] In some embodiments, the electronic device (e.g., 1400) to adjust the volume of the currently playing media comprises: the electronic device (e.g., 1400) adjusts a media playback volume (e.g., 1430 of FIG. 14M); and the electronic device (e.g., 1400) adjusts a virtual assistant volume (e.g., 1432 of FIG. 14M). In some embodiments, the virtual assistant volume is proportional to the media playback volume.

[0555] In some embodiments, the virtual assistant volume is non-zero when the media playback volume is zero.

[0556] In some embodiments, a maximum value of the virtual assistant volume (e.g., 1432) is lower than a maximum value of the media playback volume (e.g., 1430).

[0557] Note that details of the processes described above with respect to method 1500 (e.g., FIG. 15) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 1500. For brevity, these details are not repeated below.

[0558] FIGS. 16A-16P illustrate exemplary user techniques and interfaces for adding media to a playback queue, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIGS. 17-18.

[0559] FIG. 16A illustrates an exemplary scenario in which a plurality of devices are connected. Specifically, device 1600 is a smart speaker that is connected to both devices 1602 and 1604 (e.g., personal devices). In some embodiments, device 1600 includes one or more features of devices 100, 300, 500, or 580. In some embodiments, devices 1602 and 1604 include one or more features of devices 100, 300, 500, or 580. In some examples, the respective users of devices 1602 and 1604 may desire to add media to a shared payback queue that is maintained by, and/or stored in, device 1600. FIGS. 16B-16P illustrate exemplary techniques for doing so.

[0560] At FIG. 16B, device 1602 displays an exemplary playlist selection interface 1606 on display 1603. Under interface heading 1606A, device 1602 displays a playlist affordance 1606B, representing the playlist titled "Tim's Birthday". As shown in FIG. 16B, the playlist "Tim's Birthday" is identified as a "Shared Playlist". In some embodiments, a shared queue (e.g., a shared playlist) is a queue that one or more users and/or one or more devices can access and/or manage (e.g., add media items, remove media items, begin playback, or the like). At FIG. 16C, device 1602 receives user input 1608, representing selection of affordance 1606B. In some embodiments, device 1602 displays interface 1610 of FIG. 16D in response to receiving user input 1608.

[0561] FIG. 16D illustrates an exemplary interface representing a playback queue (also referred to variously as a "media playback queue" or simply a "queue"). Playlist interface 1610 of FIG. 16D includes a single media item 1610B listed under playlist identifier 1610A. As shown in FIG. 16D, media item 1610B represents the song "Same Thing" by artist "Jambug". In some embodiments, an interface representing a playback queue includes additional information associated with a media item. For example, media item 1610B also includes an identification of the user who added the media item to the playlist (e.g., the user named Tim) and a source of the media item (e.g., a media streaming service named Apple Music).

[0562] In this example, playlist interface 1610 is displayed on device 1602. However, because the playback queue represented by playlist interface 1610 is a shared queue, it can be

accessed by multiple users and/or multiple devices. In some embodiments, device 1604 displays playlist interface 1610. In some embodiments, device 1602 and device 1604 display (e.g., concurrently) playlist interface 1610. Referring back to the scenario in FIG. 16A, devices 1602 and 1604 are associated with different users.

[0563] While reference is made to a playlist of songs in this example, an exemplary playback queue is not limited to songs. In some embodiments, a playback queue is any data representing one or more media items and information usable to access the media items (e.g., for playback). In some embodiments, a playback queue includes data associated with the one or more media items. For example, the playback can include information regarding one or more of: when the media was added to the queue, when the media expires from the queue, a source of the media (e.g., file location, a URL, an identification of a streaming service, a source device (e.g., the device that stores the media item)), credentials for accessing media (e.g., login credentials or a token granting access to a media streaming service), an identifier of a user that added the media to the queue, an identifier of a device that added the media to the queue, or the like. In some embodiments, a playback queue includes, but is not limited to, one or more of the following types of media items: an audio file, a video file, an audio stream, a video stream, or the like.

[0564] Turning back to the figures, FIGS. 16E-16P illustrate exemplary interfaces for adding media items to the playlist depicted in playlist interface 1610, in accordance with some embodiments.

[0565] At FIG. 16E, device 1602 displays browsing interface 1612, which is identified by browsing interface header 1612A as a list of New Music. In some embodiments, device 1602 displays browsing interface 1612 subsequent to displaying playlist interface 1610. In some embodiments, a browsing interface is any interface (e.g., of a media application) for browsing and selecting media items. Browsing interface 1612 includes affordance 1612B representing a first media item (e.g., the song “Monday Morning” by artist “The Fomoers”) and affordance 1612C representing a second media item (e.g., the song “Paradise” by artist “MKTS”). At FIG. 16F, device receives user input 1614, associated with affordance 1612B. In some examples, user input 1614 is a deep press or a press and hold user input.

[0566] At FIG. 16G, device 1602 displays media menu 1616 in response to user input 1614. Media menu 1616 includes various items for performing actions associated with the media item of affordance 1612B. Media menu 1616 includes affordance 1616A (e.g., an "add to playlist" affordance). At FIG. 16H, device 1602 receives user input 1618 representing a selection of affordance 1616A. In some examples, user input 1618 is a tap user input.

[0567] In response to user input 1618, device 1602 displays playlist interface 1610, which has been updated to include media item 1610C (FIG. 16I). Media item 1610C represents the media item associated with media menu 1616 of FIG. 16G (e.g., the song "Monday Morning" by artist "The Fomoers"). Thus, the user of device 1602 has added a media item to the shared playlist titled "Tim's Birthday". As described above, other users and devices (e.g., device 1604) can view and add media to a shared queue, as illustrated below.

[0568] At FIG. 16J, device 1604 displays a media menu 1619 on display 1605. In some embodiments, device 1604 displays media menu 1619 in response to a user input (not illustrated) associated with affordance 1612C, of browser interface 1612 (e.g., as shown in FIG. 16E), representing the second media item (e.g., the song "Paradise" by artist "MKTS"). Media menu 1619 includes affordance 1619A (e.g., an "add to playlist" affordance). As shown in FIG. 16J, device 1604 receives user input 1620, associated with affordance 1619A. In some examples, user input 1620 is a tap user input.

[0569] In response to input 1620, device 1604 displays playlist interface 1610, which has been updated to include media item 1610D (FIG. 16K). Media item 1610D represents the media item associated with media menu 1619 of FIG. 16J (e.g., the song "Paradise" by artist "MKTS"). Thus, the user of device 1604 has added a media item to the shared playlist titled "Tim's Birthday". Accordingly, in this example, two devices associated with two different users accessed and added media to a shared queue (e.g., the media queue "Tim's Playlist", as represented by 1610).

[0570] FIG. 16K also depicts additional information regarding one or more of the media items in the playback queue represented by playlist interface 1610, in accordance with some embodiments. As shown, each affordance includes information regarding a user (e.g., a user account) that added the corresponding media item to the queue. Additionally, each affordance

includes information regarding a source (e.g., a media application) of the corresponding media item to the queue. As shown, media item 1610B was added by the user "Tim" from the source "Apple Music" (e.g., a music application, for playing local files or media streams). Media item 1610C was added by the user "Jessica" (identifier 1610G) from the source "Music Streamer" (e.g., a music streaming application) (identifier 1610E). Media item 1610D was added by the user "Jim" (identifier 1610H) from the source "Music Website" (e.g., a music streaming website) (identifier 1610F).

[0571] In some embodiments, users can remove media items from a shared queue. FIGS. 16L-16O depict exemplary interfaces for removing media items from a shared queue. At FIG. 16L, device 1604 receives user input 1621 associated with affordance 1610C of playlist interface 1610. In some examples, user input 1621 is a deep press or a press and hold user input.

[0572] In response to input 1621, device 1604 displays media menu 1622, which includes an affordance 1622A for deleting a media item from the playlist (e.g., that includes the text "Delete from Tim's Birthday") (FIG. 16M). At FIG. 16M, device 1604 receives user input 1623 representing selection of affordance 1622A. In some examples, user input 1623 is a tap user input.

[0573] At FIG. 16N, in response to input 1623, device 1604 displays playlist interface 1610, which has been updated to remove the media item associated with affordance 1610C. Thus, in the examples shown in FIG. 16N, the user of device 1604 has deleted a media item that was added by the user of device 1602. In some embodiments, any user can delete a media item from a shared queue. In some embodiments, a user can only delete media items that were added by them. In some embodiments, a user cannot delete any media items from a queue (e.g., even if added by them). In some embodiments, a user's ability to delete media items from a queue is subject to permissions. For example, a user can lack the ability (e.g., permission) to delete media items from a shared queue, but subsequently gain such ability after being granted permission (e.g., by a member of the home). In some embodiments, a user has the ability (e.g., permission) to remove media items from a queue by virtue of being a member of a set of users (e.g., a member of the home that includes device 1600 or another device that maintains or stores the queue).

[0574] FIG. 16O illustrates an exemplary message informing a user that they do not have permission to delete an item from a shared queue. Message box 1624 includes the message "You do not have permission to delete tracks from Tim's Birthday Playlist". In some embodiments, device 1604 displays message box 1624 in response to user input 1623. Thus, as alternative to the result shown in FIG. 16N, in some examples, device 1604 displays message box 1624 as shown in FIG. 16O if the user of device 1604 does not have permission to remove the media item that was added by the user of device 1602.

[0575] In some embodiments, a user or device requires permission to add a media item to a playback queue. For example, if a user lacks permission to add a media item to a shared queue, a message box similar to 1624 can be displayed that reads "You do not have permission to add tracks to Tim's Birthday Playlist". In some embodiments, permission to add media items are subject to the same conditions or approval states and levels described above with respect to permission to remove media items, and thus are not repeated.

[0576] FIG. 16P illustrates an exemplary queue removal condition, in accordance with some embodiments. Queue removal condition 1632 indicates that the media item 1610C will "expire" (e.g., be automatically removed from the queue) in 23 hours. In some embodiments, a media item in a playback queue is subject to being removed from the queue after a condition is met (e.g., or is not met), also referred to as a "queue removal condition". In some embodiments, in accordance with a determination that a queue removal condition associated with a media item is satisfied, a device (e.g., a device that maintains or stores the queue) removes the media item from the playback queue. In some embodiments, the queue removal condition is satisfied if the requester (e.g., 1602) that added the media item (e.g., 1610C) is no longer connected to a local area network that includes the device that maintains or stores the playback queue (e.g., 1600). In some embodiments, the queue removal condition is satisfied if a predetermined amount of time passes after the first media item was added to the queue. In some embodiments, the predetermined amount of time is 24 hours. In some embodiments, the queue removal condition is satisfied if the requester (e.g., user and/or device that added the media item to the queue) is no longer connected to the local area network and the device that maintains or stores the queue (e.g., 1600) is not logged in to a media streaming service (e.g., that is the source of the media item).

[0577] Notably, a fourth device 1630 displays playback interface 1610 of FIG. 16P. In some embodiments, the fourth device is associated with a user (e.g., user account) different than the users associated with devices 1600, 1602, and 1604. Thus, as described above, FIG. 16P further illustrates that a shared playback queue can be viewed by any number of devices and any number of users. In some embodiments, a device can access a shared queue when connected (e.g., directly) to the device that maintains or stores the queue (e.g., communication link 1039 of FIG. 10M between devices 1010 and 1000). In some embodiments, a device can access a shared queue when connected to common network as a device that maintains or stores the queue (e.g., via access point 1030 of FIG. 10M between devices 1010 and 1000).

[0578] In some embodiments, a device is associated with a media preference profile (also referred to as a "taste preference profile"). For example, device 1600 can be associated with (e.g., logged into) a user account associated with a third user (e.g., not the users of devices 1602 and 1604). The third user, in some examples, may not want the media that is added to the shared playback queue on their device (e.g., 1600) to affect their media preference profile. For example, the media preference profile can include data representing one or more of: particular media items associated the third user's account, media item playback history, media recommendations based on history or taste preferences, or the like. Thus, in some embodiments, media added to the playback queue on the first device (e.g., 1600) does not affect a media preference profile of a user account associated with the first device (e.g., 1600). In some examples, if the user of device 1602 is the same as the third user (e.g., whose account is associated with device 1600), device 1600 updates the media preference profile of third user based on media added by device 1602, but not based on media added by device 1604 (e.g., associated with a different user than the third user). Thus, the owner of a device can allow others to add music to a playback queue on a device associated with the owner's user account, without affect the device owner's taste preference profile.

[0579] FIG. 17 is a flow diagram illustrating a method for playing back media from a plurality of media streaming services using an electronic device in accordance with some embodiments. Method 1700 is performed at a device (e.g., 100, 300, 500, 580, 1600). Some operations in method 1700 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0580] As described below, method 1700 provides an intuitive way for playing back media from a plurality of media streaming services. The method reduces the cognitive burden on a user for playing back media from a plurality of media streaming services, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to play back media from a plurality of media streaming services faster and more efficiently conserves power and increases the time between battery charges.

[0581] The electronic device (e.g., 1600) receives (1702) a first request (e.g., represented by user input 1618 of FIG. 16H) to queue a first media item (e.g., 1612B) for playback on a set of devices (e.g., 1600). In some embodiments, the set of devices includes the first device (e.g., 1600). In some embodiments, the set of devices includes one or more devices. In some embodiments, the set of devices excludes the first device (e.g., 1600).

[0582] The electronic device (e.g., 1600) adds (1704) the first media item (e.g., 1610C as shown in FIG. 16K) to a playback queue (e.g., 1610A).

[0583] The electronic device (e.g., 1600) receives (1706) a second request (e.g., represented by user input 1620 of FIG. 16J) to queue a second media item (e.g., 1612C) for playback on the set of devices (e.g., 1600).

[0584] The electronic device (e.g., 1600) adds (1708) the second media item (e.g., 1610D as shown in FIG. 16L) to the playback queue (e.g., 1610A).

[0585] The electronic device (e.g., 1600) accesses (1710) the first media item (e.g., 1610C) from a first media streaming service (e.g., represented by 1610E of FIG. 16K).

[0586] The electronic device (e.g., 1600) initiates playback (1712) of the first media item (e.g., 1610C) on the set of devices (e.g., 1600);

[0587] The electronic device (e.g., 1600) accesses (1714) the second media (e.g., 1610D) item from a second media streaming service (e.g., represented by 1610F of FIG. 16K) different than the first media streaming service (e.g., 1610E).

[0588] The electronic device (e.g., 1600) initiates playback (1716) of the second media item (e.g., 1610D) on the set of devices (e.g., 1600).

[0589] Allowing a first media item from a first media streaming service and a second media item from a second media streaming service, different than the first media streaming service, to be added to a common playback queue, accessed, and played back allows the user to efficiently queue and manage playback of media from different sources using a single interface, thus requiring fewer user inputs to do so. Reducing the number of inputs needed to perform an operation enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0100] In some embodiments, the first request is received from a first requester (e.g., 1602). In some embodiments, a requester is a device (e.g., 1602). In some embodiments, a requester is a user account (e.g., a user account associated with identifier 1610G of FIG. 16K). In some embodiments, a requester is a user account associated with a device. In some embodiments, the first request includes identification of the first requester (e.g., represented by 1610G of FIG. 16K).

[0101] In some embodiments, the second request is received from a second requester (e.g., 1604) different than the first requester (e.g., 1602). In some embodiments, the second request includes identification of the second requester (e.g., represented by 1610H of FIG. 16L).

[0590] In some embodiments, the first requester is a second device (e.g., 1602) associated with a first user account (e.g., a user account associated with identifier 1610G). In some embodiments, the second requester is a third device (e.g., 1604) associated with a second user account (e.g., a user account associated with identifier 1610H).

[0591] In some embodiments, further in response to receiving the first request, the electronic device (e.g., 1600) determines whether the first requester (e.g., 1602) has permission to add a media item (e.g., 1610C) to the playback queue (e.g., 1610). In accordance with a determination that the first requester has permission to add a media item to the playback queue (e.g., 1610), the electronic device (e.g., 1600) adds the first media item to the playback queue (e.g., 1610 as shown in FIG. 16I). In accordance with a determination that first requester does not have

permission to add a media item to the playback queue, the electronic device (e.g., 1600) forgoes adding the first media item to the playback queue (e.g., 1610 remains as shown in FIG. 16D).

[0592] In some embodiments, to determine whether the first requester has permission to add a media item to the playback queue, the electronic device (e.g., 1600) determines whether a user account (e.g., a user account associated with identifier 1610G) associated with the first requester (e.g., 1602) has permission to add a media item (e.g., 1610C) to the playback queue (e.g., 1610). In accordance with a determination that the user account associated with the first requester has permission to add a media item to the playback queue, the electronic device (e.g., 1600) adds the first media item (e.g., 1610C) to the playback queue (e.g., 1610 as shown in FIG. 16I). In accordance with a determination that the user account associated with the first requester does not have permission to add a media item to the playback queue, the electronic device (e.g., 1600) forgoes adding the first media item to the playback queue (e.g., 1610 remains as shown in FIG. 16D).

[0593] In some embodiments, the electronic device (e.g., 1600) receives, from the second requester (e.g., 1604), a third request (e.g., 1623) to remove the first media item (e.g., 1610C) from the playback queue (e.g., 1610). The electronic device (e.g., 1600) determines whether the second requester has permission to remove the first media item from the playback queue. In accordance with a determination that the second requester has permission to remove the first media item from the playback queue, the electronic device (e.g., 1600) removes the first media item from the playback queue (e.g., 1610 as shown in FIG. 16N). In accordance with a determination that the second requester does not have permission to remove the first media item from the playback queue, the electronic device (e.g., 1600) forgoes removing the first media item from the playback queue (e.g., 1610 remains as shown in FIG. 16K).

[0594] In some embodiments, to determine whether the second requester (e.g., 1604) has permission to remove the first media item from the playback queue, the electronic device (e.g., 1600) determines whether the second requester (e.g., 1604) is a member of a predefined set of requesters (e.g., the set of users represented by 603A-603C of FIG. 6A) associated with the first device. In accordance with a determination that the second requester is a member of a predefined set of requesters associated with the first device (e.g., 1600), the electronic device

(e.g., 1600) removes the first media (e.g., 1610C) item from the playback queue (e.g., as shown in 1610 of FIG. 16N). In accordance with a determination that the second requester is a member of a predefined set of requesters associated with the first device, the electronic device (e.g., 1600) forgoes removing the first media item from the playback queue (e.g., 1610 remains as shown in FIG. 16K).

[0595] In some embodiments, the electronic device (e.g., 1600) transmits, to a fourth device (e.g., 1630 of FIG. 16P), an indication of the content of the playback queue (e.g., playlist 1610 of FIG. 16P) for display on the fourth device. In some embodiments, the indication of the content of the playback queue includes: an indication (e.g., 1610G) of the first requester that is associated with the first media item (e.g., 1610C) in the playback queue (e.g., 1610 of FIG. 16P), and an indication (e.g., 1610H) of the second requester that is associated with the second media item (e.g., 1610D) in the playback queue.

[0596] In some embodiments, subsequent to adding the first media item to the playback queue: the electronic device (e.g., 1600) determines whether a queue removal condition (e.g., illustrated by 1632 of FIG. 16P) for the first media item (e.g., 1610C) is satisfied. In some embodiments, the queue removal condition is satisfied if the requester (e.g., 1602) that added the media item (e.g., 1610C) is no longer connected to a local area network that includes the first device (e.g., 1600). In some embodiments, the queue removal condition is satisfied if a predetermined amount of time passes after the first media item was added to the queue. In some embodiments, the predetermined amount of time is 24 hours. In some embodiments, the queue removal condition is satisfied if the requester is no longer connected to the local area network and the first device (e.g., 1600) is not logged in to the first media streaming service. In some embodiments, in accordance with a determination that the queue removal condition for the first media item is satisfied, the electronic device (e.g., 1600) removes the first media item from the playback queue (e.g., removal of a media item as shown in FIG. 16O). In some embodiments, in accordance with a determination that the queue removal condition for the first media item is not satisfied, the electronic device (e.g., 1600) forgoes removing the first media item from the playback queue.

[0597] In some embodiments, the first device (e.g., 1600) is not included in the set of devices.

[0598] Note that details of the processes described above with respect to method 1700 (e.g., FIG. 17) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 1700. For brevity, these details are not repeated below.

[0599] FIG. 18 is a flow diagram illustrating a method for playing back media requested by a plurality of devices using an electronic device in accordance with some embodiments. Method 1800 is performed at a device (e.g., 100, 300, 500, 580, 1600). Some operations in method 1800 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0600] As described below, method 1800 provides an intuitive way for playing back media requested by a plurality of devices. The method reduces the cognitive burden on a user for playing back media requested by a plurality of devices, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to play back media requested by a plurality of devices faster and more efficiently conserves power and increases the time between battery charges.

[0601] The electronic device (e.g., 1600) receives (1802), from a second device (e.g., 1602) associated with a first user (e.g., a first user account), a first request (e.g., represented by user input 1618 of FIG. 16H) to queue a first media item (e.g., 1612B of FIG. 16E) for playback on a set of devices (e.g., 1600). In some embodiments, the set of devices includes the first device (e.g., 1600). In some embodiments, the set of devices includes one or more devices. In some embodiments, the set of devices excludes the first device (e.g., 1600).

[0602] The electronic device (e.g., 1600) adds (1804) the first media item (e.g., 1610C of FIG. 16K) to a playback queue (e.g., 1610 of FIG. 16K);

[0603] The electronic device (e.g., 1600) receives (1806), from a third device (e.g., 1604) associated with a second user (e.g., a second user account), a second request (represented by user

input 1620 of FIG. 16J) to queue a second media item (e.g., 1612C of FIG. 16E) for playback on the set of devices (e.g., 1600). In some embodiments, the second user is different than the first.

[0604] The electronic device (e.g., 1600) adds (1808) the second media item (e.g., 1610D of FIG. 16K) to the playback queue (e.g., 1610 of FIG. 16L).

[0605] The electronic device (e.g., 1600) initiates playback (1810) of the first media item (e.g., 1610C of FIG. 16K) on the set of devices (e.g., 1600).

[0606] The electronic device (e.g., 1600) initiates playback (1812) of the second media item (e.g., 1610D of FIG. 16L) on the set of devices (e.g., 1600).

[0607] Allowing a first media item from a first device associated with a first user and a second media item from a second device associated with a second user, different than the first user, to be added to a queue for playback on a common set of devices allows a plurality of users to efficiently share and manage a common playback queue from any of a plurality of devices, thus requiring fewer user inputs to do so. Reducing the number of inputs needed to perform an operation enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0608] In some embodiments, the first request includes a first identifier (e.g., 1610G of FIG. 16K) of at least one of the first user and the second device.

[0609] In some embodiments, the second request includes a second identifier (e.g., 1610H of FIG. 16K) of at least one of the second user and the third device.

[0610] In some embodiments, the second device (e.g., 1602) is associated with a first user account that is associated with the first user. In some embodiments, the first user account is a user account on a media streaming service. For example, exemplary media streaming services include Apple Music, Spotify, or the like. In some embodiments, the first user account is a cloud-based service user account. For example, exemplary media streaming services include Apple iCloud, or the like.

[0611] In some embodiments, the third device (e.g., 1604) is associated with a second user account that is associated with the second user.

[0612] In some embodiments, the electronic device transmits, to a fourth device (e.g., 1630 of FIG. 16P), an indication (e.g., playlist 1610 of FIG. 16P) of the content of the playback queue for display on the fourth device, wherein the indication of the content of the playback queue includes: the first identifier (e.g., 1610G or 1610E of FIG. 16P) associated with the first media item (e.g., 1610C of FIG. 16P) in the playback queue, and the second identifier (e.g., 1610H or 1610F of FIG. 16P) associated with the second media item (e.g., 1610D of FIG. 16P) in the playback queue.

[0613] In some embodiments, further in response to receiving the second request, and in accordance with a determination that the third device (e.g., 1604) has permission to add a media item to the playback queue, the electronic device (e.g., 1600) adds the second media (e.g., 1610D of FIG. 16K) item to the playback queue. In some embodiments, in accordance with a determination that the third device does not have permission to add a media item to the playback queue, the electronic device (e.g., 1600) forgoes adding the second media item to the playback queue (e.g., 1610 remains as shown in FIG. 16I).

[0614] In some embodiments, further in response to receiving the second request: in accordance with a determination that the second user (e.g., 1610H of FIG. 16K) associated with the third device (e.g., 1604) has permission to add a media item (e.g., 1610D of FIG. 16K) to the playback queue, the electronic device (e.g., 1600) adds the second media item to the playback queue (e.g., as shown in FIG. 16K). In some embodiments, in accordance with a determination that the second user associated with the third device does not have permission to add a media item to the playback queue, the electronic device (e.g., 1600) forgoes adding the second media item to the playback queue (e.g., the playback queue remains as shown in FIG. 16I, which does not include media item 1610D).

[0615] In some embodiments, the electronic device (e.g., 1600) receives, from the third device (e.g., 1604), a third request (e.g., 1623 of FIG. 16M) to remove the first media item (e.g., 1610C) from the playback queue. In some embodiments, in response to receiving the third request, and in accordance with a determination that the third device has permission to remove

the first media item from the playback queue, the electronic device (e.g., 1600) removes the first media item from the playback queue (e.g., 1610 as shown in FIG. 16N). In some embodiments, in accordance with a determination that the third device does not have permission to remove the first media item from the playback queue, the electronic device (e.g., 1600) forgoes removing the first media item from the playback queue (e.g., and displays message 1624 of FIG. 16O).

[0616] In some embodiments, in response to receiving the third request (e.g., 1623 of FIG. 16M), and in accordance with a determination that the second user is a member of a predefined set of users (e.g., the set of users represented by 603A-603C of FIG. 6A) associated with the first device (e.g., 1600), the electronic device (e.g., 1600) removes the first media item (e.g., 1610C) from the playback queue (e.g., 1610 as shown in FIG. 16N). In some embodiments, in accordance with a determination that the second user is not a member of a predefined set of users associated with the first device, the electronic device (e.g., 1600) forgoes removing the first media item from the playback queue.

[0617] In some embodiments, subsequent to adding the first media item to the playback queue, and in accordance with a determination that a queue removal condition (e.g., illustrated by 1632 of FIG. 16P) for the first media item (e.g., 1610C) is satisfied, the electronic device (e.g., 1600) removes the first media item from the playback queue (e.g., 1610 as shown in FIG. 16N). In some embodiments, the queue removal condition is satisfied if the requester (e.g., 1602) that added the media item (e.g., 1610C) is no longer connected to a local area network that includes the first device (e.g., 1600). In some embodiments, the queue removal condition is satisfied if a predetermined amount of time passes after the first media item was added to the queue. In some embodiments, the predetermined amount of time is 24 hours. In some embodiments, the queue removal condition is satisfied if the requester is no longer connected to the local area network and the first device (e.g., 1600) is not logged in to a media streaming service (e.g., that is the source of the media item). In some embodiments, in accordance with a determination that the queue removal condition for the first media item is not satisfied, the electronic device (e.g., 1600) forgoes removing the first media item from the playback queue.

[0618] In some embodiments, the first device (e.g., 1600) is not included in the set of devices.

[0619] In some embodiments, the first device (e.g., 1600) is associated with a media preference profile, wherein the media preference profile is associated with the first user. In some embodiments, further in response to receiving, from the second device (e.g., 1602) associated with the first user, the request (e.g., 1618) to queue the first media item (e.g., 1610C) for playback, the electronic device (e.g., 1600) updates the media preference profile based on the first media item. In some embodiments, further in response to receiving, from the third device associated with the second user, the request to queue the second media item for playback, the electronic device (e.g., 1600) forgoes updating the media preference profile based on the second media item. For example, if the request to queue the first media item is received from a device or user that is not associated with an owner of device 1600 (e.g., a home member), then a request to queue the media item will not affect the owner's taste preferences (e.g., which can be based on history of media played back on device 1000, or other home devices).

[0620] Note that details of the processes described above with respect to method 1800 (e.g., FIG. 18) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 1800. For brevity, these details are not repeated below.

[0621] FIGS. 19A-19P illustrate exemplary user interfaces for using a first device to configure a second device, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 20.

[0622] FIG. 19A illustrates a scenario (e.g., similar to that shown in FIG. 10E) in which a user brings device 1900 into close proximity to device 1906. In this example, device 1900 is being used to configure device 1906 (e.g., an initial configuration after powering on device 1906 for the first time). As shown, device 1906 is outputting an audio tone signal 1904 (e.g., representing sound waves). In some embodiments, device 1906 outputs an audio tone signal (e.g., 1904) in response to an indication that device 1900 is in close proximity to device 1906. For brevity, the above description regarding placement and detection of devices in close proximity (e.g., with respect to FIGS. 10A-10O) is hereby incorporated by reference, and thus not repeated.

[0623] FIG. 19B depicts an exemplary state of device 1900 immediately prior to being placed in close proximity to device 1906. As shown, device 1900 displays a home screen (e.g., home screen 605 as shown in FIG. 6B) on display 1902. At FIG. 19C, device 1900 is placed into close proximity to device 1906 (e.g., is within the proximity zone 1908). In some embodiments, proximity is detected in a manner similar to that discussed with respect to FIGS. 10C-10G. In some embodiments, in response to detecting an indication that device 1900 entered into close proximity to device 1906, device 1900 displays proximity card 1910, as shown in FIG. 19D.

[0624] FIGS. 19D-19P illustrate exemplary interfaces of a configuration process, in accordance with some embodiments. In this example, the configuration process is a process for configuring a second device using a first device. Thus, as shown, the interfaces are displayed on a first device (e.g., device 1900), and the steps of the configuration process pertain to, for example, transmitting data to the second device, selecting or entering configuration settings, and the like. In some embodiments, the steps shown in FIGS. 19D-19P can be displayed in any order (e.g., other than that shown or described herein). In some embodiments, one or more of the steps shown in FIG. 19D-19P are omitted from a configuration process. In some embodiments, one or more steps other than those shown in FIG. 19D-19P are added to a configuration process. In some embodiments, the first device and the second device are devices that include one or more features of devices 100, 300, 500, or 580.

[0625] FIG. 19D illustrates an exemplary setup card. In some embodiments, device 1900 displays setup card 1910 in response to being placed in close proximity to device 1906. Thus, a setup card is also referred to herein as a "proximity card" as described above, the description of which is hereby incorporated by reference. In some embodiments, a device continues displaying a proximity card after being removed from close proximity to another device.

[0626] Setup card 1910 includes an identification of the device to be configured, which is the device (e.g., 1906) that device 1900 was placed in close proximity to. As shown, setup card 1910 includes the title "Speaker" (e.g., device 1906 is a smart speaker), as well as an image depicting the device (e.g., in the shape of device 1906 as shown in FIG. 19A). Setup card 1910 also includes affordance 1910A. In some embodiments, device 1900 initiates a configuration

process (or otherwise proceeds to a next step in a configuration process) in response to receiving selection of affordance 1910A.

[0627] At FIG. 19E, device 1900 displays setup card 1912. In some embodiments, device 1900 displays setup card in response to user input selection (not illustrated) of affordance 1910A in FIG. 19D. Setup card 1912 depicts an exemplary location selection interface. Setup card 1912 includes a location selection affordance 1912A (e.g., which can be a scrollable list or a movable list item selector). Setup card also includes affordance 1912B, for example, for proceeding to a next step in the configuration process. In this example, the location "John's Room" is selected using affordance 1912A, and device 1900 receives user input 1913 on affordance 1912B. In some embodiments, device 1900 transmits configuration data in response to user input (e.g., 1913) associated with a setup card. In some examples, device 1900 transmits a location identifier for "John's Room" to device 1906, in response to proceeding to a next step (e.g., in response to user input 1913) while the location "John's Room" is selected. In some examples, device 1906 receives the location identifier and saves it into memory (e.g., storing it with one or more other configuration settings). In some embodiments, device 1900 transmits the configuration data to a third device (e.g., not device 1900 or 1906). For example, device 1900 can transmit configuration data to another device within the user's home (e.g., a home control hub device, such as an Apple TV by Apple Inc. of Cupertino, California USA). In some embodiments, device 1900 displays setup card 1914 of FIG. 19F in response to user input 1913.

[0628] At FIG. 19F, device 1900 displays setup card 1914. Setup card 1914 includes a prompt to the user to share settings with device 1906 ("Share Your Settings"). In some embodiments, setup card 1914 includes an identification of settings to be shared with device 1906. For example, as shown in FIG. 19F, setup card identifies a user account (e.g., login credentials) and a home wireless network (e.g., connection credentials) to be transmitted to device 1906. In some embodiments, device 1900 is already connected to or logged into (or otherwise able to log into (e.g., the credentials are stored in memory)) the accounts or networks identified by setup card 1914. Thus, setup card 1914 provides an easy interface for identifying and permitting the sharing of certain settings with device 1906. In some embodiments, setup card 1914 allows selection of one or more of the settings to share with device 1906. For example, setup card 1914 can include selection affordances that allow selection of less than all of

the accounts listed therein, thus allowing the user to share some (but not all) settings with device 1906. As shown in FIG. 19F, the settings to be shared in this example are: credentials for a user account (e.g., associated with the "Apple ID" "JAppleseed@icloud.com") and credentials for a wireless network (e.g., for the Wi-Fi network named "Appleseed Home"). In some embodiments, device 1900 transmits the settings information to device 1900 in response to user input selection of affordance 1914B.

[0629] At FIG. 19G, device 1900 displays setup card 1916. In some embodiments, device 1900 displays setup card 1916 in response to user input selection of affordance 1914B. In some examples, setup card 1916 is displayed while device 1900 exchanges data with device 1906 (e.g., device 1900 transmits settings data as described above). For example, setup card 1916 includes status 1916A showing that the devices are connecting (or otherwise exchanging data).

[0630] At FIG. 19H, device 1900 displays setup card 1918. In some embodiments, device 1900 displays setup card 1918 subsequent to setup card 1916. For example, device 1900 displays setup card 1916 after transmitting the settings data to device 1906. Setup card 1918 includes a message 1918A. In some embodiments, the message conveys a current state of the configuration process. For example, as shown in FIG. 19H, message 1918A says "Playing Tone". Thus, in the example depicted, device 1906 is currently playing a tone (e.g., an audio tone signal). In some embodiments, while displaying message 1916A, device 1900 is listening for the tone played by device 1906. For brevity, the description of audio tone signals above (e.g., with respect to FIGS. 10A-10O) are hereby incorporate by reference.

[0631] In some embodiments, device 1900 displays a message (not illustrated) indicating that the tone was not detected by device 1900. For example, device 1900 can update message 1916A to read "Your device is listening but cannot detect a tone, please move your phone closer".

[0632] In some embodiments, setup card 1918 includes one or more fields for entering a passcode (not illustrated). For example, as described above (e.g., with respect to FIGS. 10A-10O), device 1906 can output a dictated passcode (e.g., if a tone cannot be detected by device 1900 after a threshold number of attempts). Thus, using the one or more fields of setup card 1918, a user can enter the passcode that was dictated by device 1906, for example, in order to progress to a next step in the configuration process (e.g., complete the setup process).

[0633] At FIG. 19I, device 1900 displays setup card 1920, which depicts a completion setup card. In some embodiments, device 1900 displays setup card 1920 after transmitting the settings data to device 1906. In some embodiments, device 1900 displays setup card 1920 after detecting an audio tone output by device 1906. Selection of the done affordance 1920A, in some examples, dismisses the setup card 1920.

[0634] As described previously, a device (e.g., device 1906) can be configured to be part of a device group (e.g., a stereo pair). FIGS. 19J-19L illustrate exemplary interfaces for configuring a device to be part of a device group, in accordance with some embodiments.

[0635] At FIG. 19J, device 1900 displays setup card 1922. In some embodiments, device 1900 displays setup card 1922 in response to user input 1913 (e.g., representing selection of affordance 1912B of FIG. 19E, setting the location of device 1906). In the example shown in FIG. 19J, the location "John's Room" already includes an existing device (e.g., another smart speaker like device 1906), and the user is presented the option to create a device group that includes the existing device and device 1906. The user can select either affordance 1922A ("Use as a stereo pair") or affordance 1922B ("Not now"). In some examples, selection of affordance 1922B continues the configuration process as described above (e.g., display of setup card 1914). As shown in FIG. 19J, device 1900 receives user input 1923 representing selection of affordance 1922A. In some embodiments, device 1900 displays setup card 1924 in response to receiving user input 1923.

[0636] FIG. 19K illustrates an exemplary device group configuration interface. Setup card 1924 includes affordances 1924A and 1924B for selecting an audio channel for the device being configured (e.g., device 1906). In this example, the device group being created is a stereo pair (e.g., two devices that will be used to respectively output left and right channel stereo audio). In some embodiments, selection of affordance 1924C cancels the creation of a device group, and continues the configuration process as described above (e.g., display of setup card 1914). At FIG. 19K, device 1900 receives user input 1925 representing selection of the left channel (denoted "L"). Thus, device 1906 will be configured to output the left channel of a stereo audio signal for the device group named "John's Room Speakers".

[0637] In some embodiments, device 1900 displays setup card 1926 of FIG. 19L in response to receiving user input1925. FIG. 19L indicates that the creation of the stereo pair was successful and includes affordance 1926A. In some embodiments, selection of affordance 1926A continues the configuration process as described above (e.g., display of setup card 1914).

[0638] FIGS. 19M-19N illustrate exemplary interfaces for creating a new location. In some examples, the configuration process allows for a user to create a new location identifier to associate with device 1906. Setup card 1928 of FIG. 19M illustrates an alternative version of the location selection setup card 1912 of FIG. 19E. Setup card 1928 includes an option, selected by affordance 1928A, to create a new location (e.g., as indicated by the text "Create New Room"). FIG. 19N illustrates an exemplary interface 1930 for entering a location identifier. In some embodiments, interface 1930 is displayed as a proximity card. Interface 1930 includes a field 1930A for entering a name of a new location (e.g., room), for example, using displayed keyboard 1930B. As shown, the user of device 1900 has entered the name "Tim's Room" into field 1930A. Device 1900 receives user input 1931 representing selection of a done affordance. In some embodiments, device 1900 creates the location named "Tim's Room" and associates device 1906 therewith (e.g., transmits the identifier to device 1900) in response to selection of the done affordance. Thus, in some examples, the user is not required to exit, or otherwise terminate, the configuration process in order to seek out an interface (e.g., an interface of a home control application, as described above with respect to FIG. 6A) for creating a new location identifier for their home group of devices, only to return and restart the configuration process from the beginning, wasting time and device resources. In some embodiments, in response to receiving user input 1931, device displays setup card 1914. In some embodiments, in response to receiving user input 1931, device 1900 proceeds to a next step in a configuration process.

[0639] In some embodiments, a location identifier is a name associated with the device. For instance, in some examples, a location identifier does not include reference to any particular location (e.g., "Speaker 1"). Thus, the location identifier can simply be used as a name by which to identify a configured device (e.g., at a multi-device interface as described above).

[0640] In some examples, the configuration process requires that device 1900 logs into a user account in order to proceed with the configuration process. FIGS. 19O-19P illustrate exemplary

interfaces for logging into a user account during a configuration process, in accordance with some embodiments. Setup card 1932 includes the message "Please login to your iCloud account to continue", and includes an affordance 1932A that includes the option "Login Now". In some embodiments, setup card 1932 is displayed prior to setup card 1914. For example, if device 1900 determines that it is not logged into a required user account, it displays setup card 1932 in response. In some embodiments, after receiving login credentials, device 1900 displays setup card 1914. In some embodiments, device 1900 displays setup card 1932 in response to selection of affordance 1914B ("Share Settings").

[0641] At FIG. 19P, device 1900 displays a setup card 1934 with fields 1934A and 1934B for receiving login credentials during the configuration process. Thus, in some examples, the user is not required to exit, or otherwise terminate, the configuration process in order to seek out a login page, only to return and restart the configuration process from the beginning, wasting time and device resources. In some embodiments, after receiving login credentials in fields 1934A and 1934B (e.g., and after verifying them or using them to successfully login to the user account), device displays setup card 1914. In some embodiments, after receiving valid credentials (e.g., via setup card 1934), device 1900 proceeds to a next step in a configuration process.

[0642] FIG. 20 is a flow diagram illustrating a method for configuring a second device using an electronic device in accordance with some embodiments. Method 2000 is performed at a device (e.g., 100, 300, 500, 580, 1900) with a display. Some operations in method 2000 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0643] As described below, method 2000 provides an intuitive way for configuring a second device. The method reduces the cognitive burden on a user for configuring a second device, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to configure a second device faster and more efficiently conserves power and increases the time between battery charges.

[0644] The electronic device (e.g., 1900) (also referred to as a "first device") detects (2002) an audio tone signal (e.g., 1904) indicative of a pairing handshake with a second device (e.g.,

1906). In some embodiments, the audio tone signal is an initial tone. For example the tone is played in response to the devices being place in close proximity. In some embodiments, the audio tone signal is played during a configuration process. In some embodiments, the audio tone signal represents a response to another audio tone signal played by the electronic device.

[0645] In some embodiments, the first device (e.g., 1900) is connected (2004) to the second device via a communication link (e.g., 1039 of FIG. 10M). In some embodiments, the communication link is a wireless communication link. For example, the communication can be a Bluetooth or a Wi-Fi connection, or the like.

[0646] In some embodiments, the audio tone signal (e.g., 1904) is out-of-band to the communication link (2006). For example, the audio tone signal is not transmitted using the communication link. For example, the first and second devices are connected and exchange data communications over a Bluetooth wireless connection (e.g., a communication link). In this example, while the devices continue to be connected via the Bluetooth connection (e.g., while the devices are paired), an audio tone signal (e.g., an audible tone that includes or represents a communication between the devices) is out-of-band with the Bluetooth communication link. In some embodiments, being out-of-band to a communication link means that a signal (e.g., an audio tone signal) is not communicated using the same data exchange technique.

[0647] Using an audio tone signal that is out-of-band with a communication link between two devices allows the user to efficiently move through a configuration process for a device with fewer required user inputs, while ensuring that the devices are in relatively close proximity and that the user has physical access to both devices. Providing an optimized configuration process to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0648] In accordance with detecting the audio tone signal, the first device (e.g., 1900) initiates (2008), on the first device (e.g., 1900), a configuration process for configuring the second device (e.g., 1906).

[0649] Using a detection of an audio tone signal allows the user to efficiently move through a configuration process for a device with fewer required user inputs. Providing an optimized configuration process to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0650] During the configuration process, the first device (e.g., 1900) displays (2010), on the display (e.g., 1902), a configuration user interface (e.g., 1912) that includes a prompt to select a location (e.g., "Where is this speaker?" as shown in FIG. 19E).

[0651] The first device (e.g., 1900) receives (2012) user input selection (e.g., 1913) of a location (e.g., the location "John's Room" shown in 1912A of FIG. 8J).

[0652] In response to receiving the user input, the first device (e.g., 1900) configures (2014) the second device (e.g., 1906) to be associated with the location. In some embodiments, configuring the second device to be associated with the location includes transmitting configuration data that includes the location. In some embodiments, the data is transmitted to the second device (e.g., a smart speaker). In some embodiments, the data is transmitted to a third device (e.g., Apple TV).

[0653] In some embodiments, the audio tone signal (e.g., 1904) is generated by the second device (e.g., 1906 as shown in FIG. 19A).

[0654] In some embodiments, the audio tone signal includes an audio passcode. In some embodiments, the audio passcode is dictated (e.g., as shown in FIG. 10J). In some embodiments, the audio passcode is encoded in the audio tone signal.

[0655] In some embodiments, the audio tone signal (e.g., 1904) is received while the first device (e.g., 1900) is physically positioned within a threshold physical proximity (e.g., 1908 of FIG. 19C) to the second device (e.g., as shown in FIG. 19C).

[0656] In some embodiments, the audio tone signal is a first audio tone signal, the method further comprising, during the configuration process (e.g., while displaying setup card 1918 of FIG. 19H): detecting a second audio tone signal comprising encoded data. For example, the second audio tone signal includes one or more of user account login information, an identifier for the second device, or other data for confirming that the second device should be provided access to the user's account or personal data. In response to detecting the second audio tone signal: the first device (e.g., 1900) transmits an acknowledgement to the second device (e.g., 1906); and the first device (e.g., 1900) proceeds to a next step in the configuration process (e.g., displays the setup card 1920 of FIG. 19I).

[0657] In some embodiments, further in response to receiving the user input selection (e.g., 1913) of the location (e.g., "John's Room" as shown in FIG. 19J): in accordance with a determination that a third device (e.g., device 1020 of FIG. 10K, where the second device is device 1000 of FIG. 10K) is currently associated with the location (e.g., "John's Room"), the first device (e.g., 1900) displays, on the display (e.g., 1902), a configuration user interface (e.g., setup card 1922 of FIG. 19J) that includes a prompt (e.g., affordance 1922A in FIG. 19J that includes text "Use as a stereo pair") to select whether to create a device group (e.g., stereo pair) that includes the second device and the third device. In some embodiments, a device group includes two or more devices. In accordance with a determination that a third device is not currently associated with the location, the first device (e.g., 1900) forgoes displaying, on the display, the configuration user interface that includes the prompt to select whether to create the device group (e.g., displays setup card 1914 of FIG. 19F instead of setup card 1922).

[0658] In some embodiments, the first device (e.g., 1900) receives user input (e.g., 1923) representing an affirmative request create the device group. In some embodiments, in response to the user input representing an affirmative request create the device group, the first device (e.g., 1900) displays, on the display, a configuration user interface (e.g., 1924 of FIG. 19K) that includes a prompt (e.g., "Is this speaker left or right?" as shown in FIG. 19K) to select an audio output channel (e.g., left or right, as shown in FIG. 19K) to associate with the second device. In some embodiments, the first device (e.g., 1900) receives user input (e.g., 1925) representing selection of an audio output channel (e.g., left, as shown in FIG. 19K). In response to receiving the user input representing selection of a selected audio output channel, the first device (e.g.,

1900) associates (e.g., transmits configuration data to) the second device with the selected audio channel.

[0659] In some embodiments, the prompt to select a location includes a plurality of predetermined location identifiers (e.g., Kitchen, Bedroom, Patio, Dining Room, as shown in FIG. 19E).

[0660] In some embodiments, the plurality of predetermined location identifiers includes a user-defined location identifier (e.g., John's Room, as shown in FIG. 19E).

[0661] In some embodiments, the prompt to select a location includes a field (e.g., "Create New Room" option as shown in FIG. 19M) for creating a new location identifier. In some embodiments, the first device (e.g., 1900) receives user input (e.g., 1931) associated with the field for creating a new location identifier (e.g., text entered into field 1930A of FIG. 19N), wherein the user input includes a location identifier (e.g., "Tim's Room" a shown in FIG. 19N). The first device (e.g., 1900) creates a new location identifier based on the received user input associated with the field for creating the new location identifier, and associates the second device with the new location identifier (e.g., transmits configuration data that includes the new location identifier to the second device).

[0662] In some embodiments, during the configuration process: the first device (e.g., 1900) displays, on the display (e.g., 1902), a configuration user interface (e.g., setup card 1914 of FIG. 19F) that includes a prompt (e.g., "Share your settings" affordance 1914B as shown in FIG. 19F) requesting permission to transmit user account information (e.g., login credentials) to the second device (e.g., 1906). The first device (e.g., 1900) receives user input (e.g., selection of affordance 1914B) providing permission to transmit the user account information to the second device. In response to receiving the user input providing permission, the first device (e.g., 1900) transmits the user account information to the second device (e.g., 1906).

[0663] In some embodiments, prior to transmitting the user account information to the second device, and in accordance with a determination that the first device (e.g., 1900) is not currently logged into a user account associated with the user account information (e.g., is not currently logged into an iCloud account), the first device (e.g., 1900) displays, on the display

(e.g., 1902), a prompt to log in to the user account (e.g., setup card 1932 that includes the text "Please login to your iCloud account to continue" as shown in FIG. 19O), and forgoes transmitting the user account information to the second device. In accordance with a determination that the first device is currently logged into the user account associated with the user account information, the first device (e.g., 1900) transmitting the user account information to the second device (e.g., 1906).

[0664] In some embodiments, in response to progressing from a first step (e.g., setup card 1912) in the configuration process to a second step (e.g., setup card 1914) in the configuration process, the first device (e.g., 1900) outputs a third audio tone signal (e.g., 1044 of FIG. 10N). In some embodiments, the third audio tone signal is outputted concurrently with the output of a fourth audio tone signal (e.g., 1042 of FIG. 10N) outputted by the second device (e.g., 1906). In some embodiments, the third audio tone signal and the fourth audio tone signal are harmonics.

[0665] In some embodiments, in response to progressing from the second step (e.g., setup card 1914) in the configuration process to a third step (e.g., setup card 1916) in the configuration process, the first device (e.g., 1900) outputs a fifth audio tone signal (e.g., 1048 of FIG. 10O). In some embodiments, the fifth audio tone signal is outputted concurrently with the output of a sixth audio tone signal (e.g., 1046 of FIG. 10O) outputted by the second device (e.g., 1906). In some embodiments, the fifth audio tone signal and the sixth audio tone signal are harmonics. In some embodiments, the fifth audio tone signal is a higher frequency tone than the third audio tone signal (e.g., as shown in plot 1040 of FIG. 10O).

[0666] Note that details of the processes described above with respect to method 2000 (e.g., FIG. 20) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 2000. For brevity, these details are not repeated below.

[0667] FIGS. 21A-21I illustrate exemplary user interfaces for adjusting a device output of an audible signal, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 22.

[0668] FIG. 21A illustrates an exemplary scenario in which a plurality of devices are connected. Specifically, device 2100 (also referred to as a first device) is a smart speaker that is connected to device 2110 (e.g., another smart speaker, also referred to as a second device). Device 2100 is optionally connected to the device being held by user 2106 (e.g., a personal device). In some embodiments, devices 2100, 2107, and 2110 include one or more features of devices 100, 300, 500, or 580. In some examples, device 2100 and/or device 2110 includes one or more audio signal detection components (e.g., microphones) and can detect voice user input. As shown in FIG. 21A, the user 2106 utters the voice input command 2108 "Hey Siri" aloud, which is detected by device 2100. Further, the voice input command 2108 is uttered while device 2100 is currently outputting an audible signal 2104 (e.g., media playback, voice assistant text-to-speech output), and while device 2110 is currently outputting an audible signal 2112 (e.g., media playback, voice assistant voice output). In some embodiments, device 2100 and device 2110 are concurrently outputting synchronized audible signals (e.g., are a stereo pair group, and are outputting synchronized playback of the same media item).

[0669] FIG. 21B depicts an overhead view of the scenario of FIG. 21A, and illustrates the relative spatial positioning of device 2100, device 2110, and user 2106 relative to each other.

[0670] In some embodiments, in response to detecting the user input voice command (e.g., 2108 by user 2106), device 2100 adjusts the output of an audible signal (e.g., 2104). In some embodiments, device 2100 adjusts the output of the audible signal based on a spatial position of the voice command source (e.g., the position of user 2106) while continuing to output the audible signal (e.g., 2104). For example, device 2100 applies spatial domain audio processing to project the sound away from the user 2106, to filter the sound with low pass, and/or band pass, and/or high pass filters, or to lower the volume (e.g., to zero) of one or more of its speakers (e.g., where device 2100 includes an array of a plurality of speakers). For example, device 2100 can lower the volume of a speaker (e.g. of an array of speakers of device 2100) that is outputting the audible signal and that is nearest the spatial position of user 2106.

[0671] In some embodiments, in response to detecting the user input voice command (e.g., 2108), device 2100 determines a spatial position of a voice command source (e.g., user 2106). FIG. 21C illustrates exemplary values that a device (e.g., 2100) uses to determine a spatial

position of a voice command source (e.g., 2106). As shown, user 2106 is standing at a position that is a lateral angle 2122 (e.g., θ (theta) degrees) from a reference point on reference marker 2120. Reference marker 2120 is included for illustrative purposes. User 2106 is standing at a distance 2124 (e.g., d units of length) away from device 2100 (e.g., which is depicted as the center of reference marker 2120). In some embodiments, device 2100 uses one or more of an angle and distance to determine a spatial position of a voice command source. In some embodiments, an angle is a lateral angle (e.g. as shown in FIG. 21C). In some embodiments, an angle is a vertical angle. For example, using a vertical angle (not illustrated), device 2100 can determine a height (e.g., in a z-direction, a third dimension) of a voice command source relative to device 2100. In some embodiments, a spatial position of the voice command source is a position relative to device 2100 (e.g., relative to a coordinate system of device 2100, as illustrated by 2120).

[0672] In some embodiments, a plurality of devices (e.g., 2100 and 2110) are used to determine a spatial position of a voice command source (e.g., user 2106). For example, both devices 2100 and 2110, as shown in FIG. 21C, can determine a distance to user 2106 and share the results with the other device. In some examples, device 2100 uses both distance measurements to determine a spatial position of user 2106. For further example, three devices can be used to triangulate a position of a voice command source (e.g., user 2106). Utilizing a plurality of approximations or measurements from a plurality of devices, in some examples, results in a more accurate determination of spatial position of a voice command source than would otherwise be achieved with a single device.

[0673] In some embodiments, device 2100 adjusts the output of the audible signal based on the spatial position of the voice command source (e.g., user 2106). FIG. 21D illustrates the overhead view of FIGS. 21C and 21D, however the output of the audible signal 2104 has been adjusted (e.g., reduced) in the direction of user 2106, as shown by the lack of sound waves emanating from device 2100 in the direction of user 2106. As seen in FIG. 21D, device 2100 continues to output audible signal 2104 in directions not directly facing the user.

[0674] In some embodiments, device 2100 adjusts the output of the audible signal based on the second device (e.g., 2110). FIG. 21E illustrates that device 2100 has adjusted (e.g., ceased)

the output of the audible signal 2104 (similar to FIG. 21D) both in the direction of user 2106 and in the direction of the second device 2110. This can be done, for example, when device 2110 is also listening to the voice command input. In some examples, devices 2100 and 2110 can both detect voice user input 2108. In such examples, device 2110 can better suited (e.g., is a shorter distance from the user, has a more sensitive microphone, has faster speech processing hardware) for detecting the voice user input (e.g., 2108). Thus, for example, device 2100 can reduce the output of an audible signal so as to reduce an influence on the device 2110 (e.g., thus reducing interference with device 2110's audio detection function). In some embodiments, the first device (e.g., device 2100) uses a known influence (e.g., that its audio output has) on the second device (e.g., device 211) to adjust output. In some embodiments, the first device (e.g., device 2100) determines a known influence based on data received from the second device (e.g., device 2110). For example, device 2100 can receive (e.g., from device 2110) data representing an influence that audio output from device 2100 has on detection circuitry (e.g., a microphone array) of device 2110. In some embodiments, a known influence is determined during playback (e.g., device 2100 determines a known influence while outputting the audible signal 2104). For instance, in some examples, device 2100 and device 2110 are outputting the same output signal, and device 2100 thus uses the known output signal to reduce interference with any input signal (e.g., voice input) from user 2106 (e.g., device 2100 subtracts one signal from the other before signal output). In some embodiments, a known influence is predetermined data retrieved from memory (e.g., previously determined and subsequently retrieved from memory of device 2100). In some embodiments, a known influence (e.g., on a second device) includes data regarding the influence of one or more of the following characteristics of audio output by the first device: frequency, volume, and direction.

[0675] FIGS. 21F and 21G illustrate an exemplary adjustment of output of device 2100. As shown, device 2100 includes a speaker array (e.g., one or more speakers) of three speakers (e.g., 2102A-2102C), each pointing in different directions (e.g., directions 120 degrees apart). In some embodiments, a speaker array includes any number of two or more speakers. Before adjustment (e.g., before detection of voice user input 2108), the output levels of the speakers 2102A-2102C are shown in FIG. 21F. As shown in FIG. 21G, in response to detecting voice user input 2108, device 2100 reduces the output of speaker 2102B (e.g., which faces the user 2106) to zero (e.g., zero volume), and slightly reduces the output of speakers 2102A and 2102C.

[0676] In some embodiments, a first device performs an audio processing technique. For example, device 2100 can perform an audio processing technique to project the audio output so that it is perceived by a user to be in a direction away from the user. For example, device 2100 performs a spatial audio processing technique on one or more signals representing the audio output. The spatial audio processing technique, in some examples, is used to transform the one or more signals representing the audio output, when outputted by the device, is perceived by a listener to emanate from a selected spatial position. One of skill in the art would recognize that various appropriate techniques can be used for performing such audio processing, all of which are intended to be within the scope of this disclosure.

[0677] FIG. 21H illustrates the overhead view of devices 2100 and 2110 and user 2106, and includes two points—a first point 2132 (e.g., representing a spatial position) that is in a direction away from user 2106 (e.g., relative to device 2100), and a second point 2130 (e.g., representing a spatial position) that is in a direction toward the user (e.g., relative to device 2100).

[0678] In some embodiments, device 2100 adjusts audio output such that it will be perceived by a voice command source (e.g., user 2106) to emanate from a point in a direction away from them (e.g., relative to device 2100). For example, in response to detecting voice user input 2108, the device can adjust the output of 2104 (e.g., playback of music as shown in FIG. 21I) so that the user 2106 perceives the sound to emanate from point 2132, rather than from the position of device 2100. Thus, while user is speaking, for example, they are provided an indication that device 2100 is currently listening (e.g., even though it continues playing back media) when they perceive the volume of audio output to be lowered (e.g., in comparison to the audible output prior to adjustment, as seen in FIG. 21B). Further, in this example, media playback is not interrupted (e.g., paused, stopped, muted) by a user voice command.

[0679] In some embodiments, device 2100 adjusts audio output such that it will be perceived by a voice command source (e.g., user 2106) to emanate from a point in a direction toward them (e.g., relative to device 2100). For example, as shown in FIG. 21I, in response to detecting voice user input 2108 (e.g., a command to a voice assistant function on device 2100), device 2100 outputs a response 2140 (e.g., text-to-speech output "OK, I'll set a timer for 30 minutes") in a direction toward user 2106 (e.g., toward point 2130 of FIG. 21H). In this example, voice user

input 2108 included the command "Hey Siri, set a timer for 30 minutes". Thus, user 2106 perceives the response sound (e.g., 2140) to emanate from point 2130. Thus, device 2100 utilizes a known spatial position of a voice command source to target an audible response in the direction of the source. Thus, the user is provided a response that is more likely to be heard by the user. For example, if the user is in a noisy environment, targeting the audio output to the user's position can increase the likelihood that the user hears the output of device 2100. Further, in this example, media playback is not interrupted (e.g., paused, stopped, muted at least in directions facing away from the voice command source) while providing response sound 2140, since response sound 2140 is directed towards the user is therefore readily perceptible, even while media playback continues. Further, if there are multiple users in a room, device 2100's spatially-targeted output can provide increased privacy by targeting a response such that other, non-targeted users are less likely to hear.

[0680] FIG. 22 is a flow diagram illustrating a method for adjusting audible output using an electronic device in accordance with some embodiments. Method 2200 is performed at a device (e.g., 100, 300, 500, 580). Some operations in method 2200 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0681] As described below, method 2200 provides an intuitive way for adjusting audible output. The method reduces the cognitive burden on a user for adjusting audible output, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to adjust audible output faster and more efficiently conserves power and increases the time between battery charges.

[0682] While outputting an audible signal (e.g., 2104) from the electronic device (e.g., 2100), the electronic device (e.g., 2100) detects a (2202) user input voice command (e.g., 2108) from a voice command source (e.g., 2106).

[0683] In response to detecting the user input voice command, the electronic device (e.g., 2100) determines (2204) a spatial position (e.g., angle 2122 and/or distance 2124 of FIG. 21C) of the voice command source (e.g., 2106) relative to the electronic device (e.g., 2100). In some embodiments, determining a spatial position includes determining one or more of a distance (e.g., 2124 of FIG. 21C) and an angle (e.g., 2122 of FIG. 21C) to the voice command source

(e.g., a user speaking). For example, an angle can be a lateral angle or a vertical angle relative to a coordinate system of the electronic device.

[0684] While continuing to output the audible signal (e.g., 2104), the electronic device (e.g., 2100) adjusts (2206) the output (e.g., volume, frequency, and/or direction) of the audible signal based on the spatial position of the voice command source (e.g., as shown in FIG. 12D). In some embodiments, the electronic device (e.g., 2100) applies spatial domain audio processing to the audible signal. In some embodiments, the electronic device (e.g., 2100) projects the output of the audible signal (e.g., sound) in one or more directions other than a direction toward the voice command source (e.g., 2106). In some embodiments, the electronic device (e.g., 2100) projects the output in a direction away from the voice command source. In some embodiments, the electronic device (e.g., 2100) includes an array (e.g., two or more) of speakers. In some embodiments, the electronic device (e.g., 2100) adjusts the outputs of each speaker in the array of speakers. For example, the electronic device (e.g., 2100) adjusts the relative output of one or more speakers in the array to spatially-target output of the audible signal in a direction other than toward the voice command source. In some examples, the electronic device (e.g., 2100) lowers the output volume (e.g., to zero) of a speaker (e.g., of an array of speakers) that is outputting the audible signal and that is nearest to the spatial position.

[0685] Adjusting the output of an audible signal, while continuing to output it, based on a spatial position determined in response to a user input voice command provides the user with feedback that their voice command is being detected. Providing improved audible feedback to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0686] In some embodiments, the device is a first device, and the audible signal is a first audible signal. In some embodiments, the electronic device (e.g., 2100) determines that a second device (e.g., 2110) is currently outputting a second audible signal (e.g., 2112). For example, the first and second device are currently playing the same content (e.g., are a device group), or the

first device detects (e.g., via one or more microphones) that the second device is outputting audio (e.g., which can be a different audio signal).

[0687] In some embodiments, adjusting, by the first device (e.g., 2100), the output of the first audible signal is further based on a known influence that audio output by the first device has on the second device (e.g., 2110). For example, the second device may be closer to the voice command source (e.g., 2106), and thus may be the most appropriate device to listen to the user input voice command (e.g., 2108). In some embodiments, the audio output of the first device is adjusted so as to reduce interference with the second device's ability to detect and/or process the user input voice command. For example, the second device includes a microphone for detecting voice input and the first device adjusts its output so as to reduce interference with the second devices use of its microphone to detect voice user input. In some embodiments, the known influence is determined (e.g., by device 2100) based on the spatial position of the second device (e.g., relative to the position of the first device). In some embodiments, the known influence is based on data received from the second device (e.g., the second device transmits quantitative data regarding how the output of audible signals by the first device affects the audio detection circuitry (e.g., microphone) of the second device).

[0688] In some embodiments, the first device (e.g., 2100) adjusts the output (e.g., volume, frequency, and/or direction) of the audible signal based on the spatial position of second device (e.g., 2110) (e.g., as shown in FIG. 21E).

[0689] In some embodiments, the device (e.g., 2100) includes a speaker array (e.g., two or more speakers, such as speakers 2102A-2102C of FIG. 21F), wherein each speaker in the speaker array (e.g., 2102A-2102C) receives a respective output signal. In some embodiments, adjusting the output of the audible signal based comprises: performing an audio processing technique using at least: the respective output signals of each speaker in the speaker array (e.g., as illustrated in FIG. 21F), and the spatial position of the voice command source (e.g., as shown in FIG. 21C). In some embodiments, the electronic device (e.g., 2100) adjusts at least one of the respective output signals corresponding to a speaker of the speaker array (e.g., as shown in FIG. 21G, in which speaker 2102B's output volume level has been reduced).

[0690] In some embodiments, the audio processing technique includes transforming the audio output into the spatial domain and applying one or more filters (e.g., head-related transfer functions (HRTFs). For example, the filter can create a three-dimensional audio effect at a position that is laterally located in the opposite direction of the voice command source (e.g., point 2132 as shown in FIG. 21H). Thus, if a user is located at a position that is at some lateral angular position relative to the device (e.g., an angular position relative to the speaker that will be referred to as 0 degrees for reference), the speaker can apply a 3D audio effect to adjust the output of the audible signal so that the user perceives the audio to be emanating from an angular position of 180 degrees relative to the speaker (e.g., point 2132 as shown in FIG. 21H). In some embodiments, the audio processing includes a cross-talk cancellation (CTC) process. In some embodiments, the audio processing technique includes a beamforming process (e.g., using constructive and/or destructive interference of sound waves to achieve spatial selectiveness of audio output). Any technique for creating a perceived spatial audio effect to a human listener is intended to be within the scope of this disclosure.

[0691] In some embodiments, the spatial position is a first spatial position (e.g., 2130 of FIG. 21H), and wherein adjusting the output of the audible signal further comprises outputting the audible signal (e.g., 2104) such that it will be perceived, by a listener (e.g., user 2106) at the first spatial position (e.g., 2130 of FIG. 21H), to emanate from a second spatial position (e.g., 2132 of FIG. 21H) that is in a direction, relative to the device (e.g., 2100), away from the first spatial position of the voice command source. For example, device 2100 applies a 3D audio effect such that a human listener located at the user's position will perceive the output (e.g., 2104) to be emanating from the spatial position that is in a direction away from the user, relative to the device.

[0692] In some embodiments, subsequent to adjusting the output of the audible signal: the electronic device (e.g., 2100) detects that the user input voice command (e.g., 2108) from the voice command source (e.g., 2106) has ended. In response to detecting that the user input voice command from the voice command source has ended, the electronic device (e.g., 2100) ceases adjusting the output the audible signal. For example, the adjusted audio output (shown in FIG. 12G) returns to its previous unadjusted state (e.g., as shown in FIG. 21F).

[0693] In some embodiments, determining the spatial position of the voice command source relative to the device comprises: determining a distance (e.g., 2124) of the voice command source relative (e.g., 2106) to the electronic device (e.g., 2100), and determining an angular position (e.g., a lateral or vertical angle) (e.g., 2122) of the voice command source relative to the device.

[0694] In some embodiments, the audible signal (e.g., 2104) is a first audible signal, and while adjusting the output of the audible signal (e.g., 2104 as shown in FIG. 21D): the electronic device (e.g., 2100) outputs a second audio signal (e.g. 2140) based on the spatial position of the voice command source.

[0695] In some embodiments, the spatial position (e.g., position of user 2106 in FIG. 21I) is a first spatial position, and outputting the second audible signal (e.g., 2140 as shown in FIG. 21I) comprises outputting the second audible signal such that it will be perceived, by a listener at the first spatial position, to emanate from a third spatial position (e.g., point 2130) that is in a direction, relative to the device, toward the first spatial position of the voice command source.

[0696] In some embodiments, the first audible signal represents playback of a media item (e.g., a song, video audio), and wherein the second audible signal represents a virtual assistant output (e.g., a voice assistant output 2140, as shown in FIG. 21I).

[0697] In some embodiments, the electronic device (e.g., 2100) includes a plurality of microphones (e.g., a microphone array).

[0698] Note that details of the processes described above with respect to method 2200 (e.g., FIG. 22) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 2200. For brevity, these details are not repeated below.

[0699] FIGS. 23A-23K illustrate exemplary user interfaces for blended taste profile playback queues, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 24.

[0700] FIG. 23A illustrates an exemplary scenario in which a plurality of devices are connected. Specifically, device 2300 is a smart speaker that is connected to both devices 2302 and 2310 (e.g., personal devices). In some embodiments, device 2300 includes one or more features of devices 100, 300, 500, or 580. In some embodiments, devices 2302 and 2304 include one or more features of devices 100, 300, 500, or 580. In some examples, the respective users of devices 2302 and 2310 may desire to playback media based on both of their taste profiles. In some examples, devices 2300, 2302, and/or 2310 are used to create, maintain and/or store a playback queue based on a plurality of taste profiles. FIGS. 23B-23K illustrate exemplary interfaces for doing so.

[0701] FIG. 23B illustrates an exemplary playback queue interface. In this example, playlist interface 2312 represents the playlist named "Users Nearby" and includes media items 2314A-2314C. In some embodiments, a device (e.g., 2300) automatically generates (e.g., creates or populates with media items) a media playback queue based on the taste profiles of connected devices. For example, device 2300 creates (or otherwise populate with media items) the playlist "Users Nearby" based on taste profiles associated with devices (e.g., or users associated with those devices) connected to device 2300. Thus, as shown in FIG. 23B, the playlist 2312 (displayed by device 2310) is created and the media items 2314A-2314C are added to the playlist automatically in response to detecting that a single device is connected (e.g., device 2310) to device 2300 (e.g., on the same Wi-Fi network, via Bluetooth, or the like). In some embodiments, a device (e.g., 2300) automatically generates a media playback queue (e.g., the playlist named "Users Nearby" (displayed by device 2310) depicted in FIG. 23B) in response to detecting that the device (e.g., 2300) is connected to one or more other devices (e.g., 2302 or 2310).

[0702] FIGS. 23C-23D illustrate depictions of exemplary taste profiles of two users, and two exemplary combined taste profiles. In some embodiments, a device (e.g., 2300) generates a playback queue based on one or more taste profiles. In some embodiments, a device (e.g., 2300) generates a playback queue based on a combined taste profile.

[0703] FIG. 23C illustrates taste profile 2316A (of a first user) and taste profile 2316B (of a second user). Taste profile 2316A includes the media preferences: Music Genre 1, Music Genre 2, and Music Genre 3. Taste profile 2316B includes the media preferences: Music Genre 2,

Music Genre 3, and Music Genre 4. Thus, taste profiles 2316A and 2316B have some overlap in media preference. In some embodiments, a device (e.g., 2300) generates a combined taste profile based on an overlap of media preferences. For example, device 2300 generates combined taste profile 2316C, which includes the media preferences that are common to both taste profile 2316A and 2316B, namely the media preferences: Music Genre 2 and Music Genre 3. Thus, generating a playback queue using the combined taste profile 2316C will result in a playback queue that includes media items that the users associated with the taste profiles 2316A and 2316B are both likely to enjoy.

[0704] FIG. 23D illustrates taste profile 2318A (of a first user) and taste profile 2318B (of a second user). Taste profile 2318A includes the media preferences: Music Genre 1 and Music Genre 2. Taste profile 2318B includes the media preferences: Music Genre 3 and Music Genre 4. Thus, taste profiles 2318A and 2318B do not have overlap in media preference. In some embodiments, a device (e.g., 2300) generates a combined taste profile based on one or more media preferences for each taste profile associated with a detected. For example, device 2300 generates combined taste profile 2318C, which includes at least one media preference from each of taste profile 2318A and 2318B, namely the media preferences: Music Genre 1, Music Genre 2, Music Genre 3, and Music Genre 4. Thus, generating a playback queue using the combined taste profile 2318C will result in a playback queue includes media items that at least one user associated with the profiles 2318A and 2318B is likely to enjoy.

[0705] In some embodiments, a device (e.g., 2300) populates (e.g., adds media items to) an existing media playback queue in response to detecting a device associated with a taste profile. Thus, in some examples, device 2300 dynamically updates a queue based user taste profiles as more users are detected. FIG. 23E illustrates playlist 2312 (displayed by device 2310), but it has been updated to include media items 2314D and 2314E. For example, device 2300 added media items 2314D and 2314E in response to detecting connection to device 2302. In this example, device 2300 updated the combined taste profile (e.g., used to populate playlist 2312) based on the taste profile associated with device 2302.

[0706] In some embodiments, a device (e.g., 2300) removes media items from a playback queue in response to no longer detecting a device associated with a taste profile. For example,

device 2300 updates a queue to cater to the users that remain connected. For instance, in response to device 2302 no longer being detected by device 2300, device 2300 can remove media items 2314D and 2314E from playlist 2312. Thus, media playback can be tailored to users who remain present in a location (e.g., connected to device 2300) even as users come and go, without requiring excessive user inputs to do so.

[0707] In some embodiments, a device (e.g., 2300) creates a playback queue based on a plurality of taste profiles in response to user input. FIGS. 23F-23K illustrate exemplary interfaces for doing so.

[0708] FIG. 23F illustrates an exemplary playlist selection interface 2320, that includes a playlist creation affordance 2320A. At FIG. 23G, device 2310 receives user input 2322 representing selection of affordance 2320A.

[0709] In response to user input 2320A, device 2310 displays playlist creation interface 2324 (FIG. 23H). Playlist creation interface 2324 includes a title 2324A ("Dinner Party"), an affordance 2324B for making the playlist public (e.g., make it a shared playlist that can be viewable by other users, editable by other users, or the like), and an affordance 2324C for adding music to the playlist. At FIG. 23I, device 2310 receives user input 2326 representing selection of affordance 2324C.

[0710] In some embodiments, a device (e.g., 2310) displays a prompt to create a queue based on a plurality of taste profiles of detected devices in response to receiving user input (e.g., 2326). As shown in FIG. 23J, in response to receiving user input 2326, device 2310 displays message 2328, which reads: "The following users have been detected in your home: Tim, Jessica, Sophia, and Zachary. Would you like to create a playlist based on each user's tastes?" In response to selection of affordance 2328A, device 2310 generates a playback queue (e.g., such as illustrated in FIG. 23E) based on the plurality of taste profiles associated with the detected devices (e.g., corresponding to the users named Tim, Jessica, Sophia, and Zachary). Thus, as shown in the example in FIGS. 23H-23J, device 2310 determines that a user is attempting to make a playlist and displays a prompt to create a queue based on the taste profiles of detected users.

[0711] As described above, a device can automatically generate a playback queue based on a plurality of taste profiles. In some embodiments, a device (e.g., 2300, 2310) generates the queue based on a plurality of taste profile in response to user input activating (e.g., opening) a media application. For example, FIG. 23K illustrates playlist selection interface 2320 (similar to FIG. 23F), but it has been updated to include an affordance 2320B representing a playlist that was automatically generated based on the taste profiles of detected users (e.g. Tim, Jessica, and 2 others). Thus, when a user opens a music application to browse playlists, a playlist based on a combined taste profile of detected users can already exist (e.g., and be displayed), reducing the number of inputs required to create and/or access such a queue.

[0712] FIG. 24 is a flow diagram illustrating a method for maintaining a queue based on multiple user taste profiles using an electronic device in accordance with some embodiments. Method 2400 is performed at a device (e.g., 100, 300, 500, 580). Some operations in method 2400 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0713] As described below, method 2400 provides an intuitive way for maintaining a queue based on multiple user taste profiles. The method reduces the cognitive burden on a user for maintaining a queue based on multiple user taste profiles, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to maintain a queue based on multiple user taste profiles faster and more efficiently conserves power and increases the time between battery charges.

[0714] The electronic device (e.g., 2300 or 2302), which is a first device, detects (2402) that a second device (e.g., 2310) associated with a second user account (e.g., an iCloud account) is connected to the first device (e.g., 2300 or 2302). For example, the second device is connected to the same Wi-Fi network, within Bluetooth range of the first device, or the like. The first device is associated (2404) with a first user account (e.g., an iCloud account). The first user account is associated (2406) with a first media preference profile (e.g., 2316A or 2318A). In some embodiments, a media preference profile includes data representing media items or information for identifying media items that represents the preferences of the owner of the first user account. In some embodiments, a preference profile includes one or more media items. For

example, the profile includes identifiers for media (e.g., songs, videos) that the first user account has saved, liked, listened to (e.g., a plurality of times), or otherwise indicates that the owner of the first user account likes a media item. In some embodiments, the preference profile includes generalized information for identifying media items. For example, generalized information can include a genre (e.g., type of music), an artist, an album (e.g., comprised of a plurality of media items). Such generalized information can be used to identify media items by matching, for example, metadata associated with media items to the generalized information. The second user account is associated (2408) with a second media preference profile (e.g., 2316B or 2318B) different than the first media preference profile.

[0715] The first device (e.g., 2300 or 2302) initiates (2410) playback of media based on the first media preference profile and the second media preference profile.

[0716] Initiating playback of media based on first and second media profiles associated with first and second devices that are connected provides users with an improved media playback feature for creating playback queues according to the tastes of multiple users whose devices are connected while reducing redundant user inputs at each device. Providing improved media playback features to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0717] In some embodiments, the first device (e.g., 2300 or 2302) determines a combined preference profile (e.g., 2316C or 2318C) based on the first media preference profile (e.g., 2316A or 2318A) and the second media preference profile (e.g., 2316B or 2318B).

[0718] In some embodiments, determining a combined preference profile based on the first media preference profile and the second media preference profile comprises: determining a media preference that is related to both the first media preference profile and the second media preference profile. In some embodiments, a media preference includes one or more of a genre of media (e.g., music), an artist, a media item (e.g., a song), or the like. In some embodiments, a media preference that is related to both is a media preference that is common to both profiles

(e.g., 2316C). In some embodiments, a media preference that is related to both is a media preference that is a subset or superset of both profiles. For example, rock and roll music is a superset of rock and roll music from the year 1987.

[0719] In some embodiments, while continuing to detect that the second device (e.g., 2310) is connected to the first device: the first device (e.g., 2300 or 2302) adds a plurality of media items (e.g., 2314D and 2314E of FIG. 23E) to a playback queue (e.g., the playlist "Users Nearby" shown in FIG. 23E) based on the first media preference profile (e.g., 2316A or 2318A) and the second media preference profile (e.g., 2316B or 2318B). In some embodiments, the first device (e.g., 2300 or 2302) initiates playback of the plurality of media items.

[0720] In some embodiments, the first device (e.g., 2300 or 2302) receives a request (e.g., selection of affordance 2314F) to initiate playback on the first device, and in response to receiving the request, initiates the playback of media based on the first media preference profile and the second media preference profile.

[0721] In some embodiments, the request (e.g., selection of 2314F) is received from the second device (e.g., 2310). For example, a guest device (e.g., not associated with the device or the home set of devices) initiates playback of a playback queue that includes a media items pertaining to a blended taste profile.

[0722] In some embodiments, the first device includes a display (e.g., 2302), and wherein the method further comprises: displaying, on the display, an affordance (e.g., a "play now" affordance (e.g., 2314F), a media item in a playlist (e.g., 2314D), a "shuffle all" affordance, or the like), and wherein receiving the request to initiate playback on the first device comprises receiving user input selection of the affordance (e.g., 2314F). In response to receiving user input selection of the affordance, the first device (e.g., 2300 or 2302) initiates the playback of media based on the first media preference profile and the second media preference profile.

[0723] In some embodiments, initiating playback of media based on the first media preference profile and the second media preference profile comprises: transmitting an instruction to a third device (e.g., 2300) to initiate playback.

[0724] In some embodiments, the first device (e.g., 2300 or 2302) receives a media item (e.g., an identifier of a media item, a link to the media item, or a copy of the media item). In response to receiving the media item: in accordance with a determination that the media item was received from the first device, the first device (e.g., 2300 or 2302) updates the first media preference profile based on the media item (e.g., 2316A or 2318A); and in accordance with a determination that the media item was received from the second device (e.g., 2310), the first device (e.g., 2300 or 2302) forgoes updating the first media preference profile based on the media item. For example, if the media item was added to a blended taste profile queue by a device not associated with a user account of the device managing or storing the queue (e.g., the owner's device), the owner's taste profile is not updated based on the addition of that media item to the queue.

[0725] Note that details of the processes described above with respect to method 2400 (e.g., FIG. 24) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 2400. For brevity, these details are not repeated below.

[0726] FIGS. 25A-25I illustrate exemplary user interfaces for allowing guest playback on a set of devices, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 26.

[0727] FIG. 25A illustrates an exemplary scenario in which a plurality of devices are connected. Specifically, device 2500 is a smart speaker that is connected to both devices 2502 and 2504 (e.g., personal devices). In some embodiments, device 2500 includes one or more features of devices 100, 300, 500, or 580. In some embodiments, devices 2502 and 2504 include one or more features of devices 100, 300, 500, or 580. In this example, the user (also referred to as a first user) of device 2504 is associated with a user account that is associated with device 2500 (e.g., both device 2504 and 2500 are members of a home set of devices, as described above with respect to FIG. 6A). Further, in this example, the user of device 2502 is not associated with device 2500 (e.g., is a guest of (e.g., not a member of) the home that includes the home set of devices). The user (also referred to as a second user) of device 2502 may desire access to device

2500, for example, to initiate playback therefrom. FIGS. 25B-25I illustrate exemplary interfaces for requesting and granting access, in accordance with some embodiments.

[0728] FIG. 25B illustrates an exemplary current media interface. As shown, device 2502 (e.g., the guest device) displays current media interface 2506 on display 2503. As shown, device 2502 is currently playing back a media item ("Same Thing") on itself (e.g., as indicated by device selection affordance 2506A, which identifies device 2502, named "iPhone"). At FIG. 25C, device 2502 receives user input 2508 representing selection of device selection affordance 2506A.

[0729] In response to user input 2508, device 2502 displays exemplary multi-device interface 2510 (FIG. 25D). In some embodiments, multi-device interface 2510 is a multi-device interface such as multi-device interface 610 of FIGS. 6H-6N and multi-device interface 810 of FIGS. 8C-8P, discussed above. In this example, the devices listed in interface 2510 (except 2510A, "iPhone", which represents device 2502) are devices associated with the first user (e.g., associated with device 2504 and device 2500) that are members of the home set of devices. At FIG. 25D, device 2502 receives user input 2512 representing selection of indicator 2510C, which is associated with indicator 2510B, representing the device named "Living Room" (e.g., device 2500). In some embodiments, device 2502 transmits a request to initiate playback on the selected set of devices (e.g., transmits a request to 2500).

[0730] In response to receiving selection of the device named "Tim's Living Room" (e.g., user input 2512), device 2502 displays message 2514 (FIG. 25E), which prompts the user to "Please wait". In some examples, message 2514 is displayed while permission for device 2502 is being requested (e.g., at another device). For example, subsequent to transmitting a request to initiate playback to device 2500 (e.g., and while device 2502 is awaiting confirmation that playback has been initiated or permission has not been granted), device 2502 displays message 2514. In some embodiments, the ability of device 2502 to initiate playback on a device with which it is not associated (e.g., device 2502 is not associated with same user account, or is not a member of a common set of devices (e.g., home set devices) with device 2500) is subject to permission from another device associated with device 2500. In some embodiments, the requesting device (e.g., 2502) transmits a request for permission to a device (e.g., device 2500 or

device 2504) of the set of devices associated with (e.g., that includes) device 2500. For example, in response to a user input request to initiate playback on a device of a home set of devices, device 2502 transmits a request for permission to one or more devices in the home set of devices (e.g., device 2500, or another device in the set of devices). In some embodiments, the request is transmitted by the requesting device (e.g., 2500) to another device (e.g., with a display) of the set of devices (e.g., device 2504). For example, device 2500 receives the playback request from device 2502 and, in response, transmits the request for permission (e.g., for device 2502) to device 2504.

[0731] FIGS. 25F-25G illustrate exemplary permission request interfaces. In some embodiments, a permission request interface (e.g., 2516 or 2518) is displayed on a device (e.g., by device 2504) of the set of devices (e.g., 2500 and 2504) in response to a request to initiate playback on a device (e.g., 2500) of the set of devices by a device (e.g., 2502) that is not included in the set of devices.

[0732] At FIG. 25F, device 2504 (e.g., which is included in the home set of devices) displays permission request interface 2516, which includes affordances for granting permission (2516A) and denying permission (2516B). For example, user input associated with 2516A will grant device 2502 permission to initiate media playback of "Same Thing" on device 2500, and user input user input associated with 2516B will deny device 2502 such permission. In some embodiments, a permission request interface includes information about the request to initiate playback on the set of devices. In some embodiments, the information includes one or more of, but is not limited to, an identification of: the device requesting initiation of playback (e.g., 2502), a user (e.g., a user account) associated with the request (e.g., the user of 2502), and an identification of the devices on which playback is requested (e.g., the device 2500 named Living Room). As shown in FIG. 25F, interface 2516 includes the information: "Would you like to allow Bob to play music in the Living Room?", which identifies the user (e.g., named Bob) requesting initiation of playback and the device on which they are requesting it (e.g., named or located in the Living Room).

[0733] In some embodiments, a permission interface allows selection of a plurality of different permission levels. In some embodiments, a permission level grants conditional

permission. For example, permission request interface 2518 (FIG. 25G) includes a plurality of permission levels (associated with affordances 2518A-2518D) that can be granted by the user of device 2504: "Always Allow" (e.g., always allow the user Bob to initiate playback), "Allow Once" (e.g., allow the current request, but not future requests from Bob), "Not Now" (e.g., deny the current request, but ask permission for future requests from Bob), and "Never Allow" (e.g., deny the currently request and all future requests from Bob).

[0734] FIG. 25H illustrates an updated current media interface 2506 displayed on device 2502. In some embodiments, the updated current media interface 2506 is displayed in response to device 2502 being granted permission (e.g., user input associated with affordance 2516A at device 2504). As shown, device selection affordance 2506A indicates that playback is occurring on device 2500 and device 2502 (e.g., the devices selected at FIG. 25D). In some embodiments, the current media interface includes an indication of a user associated with the device of the set of devices. For example, as shown in FIG. 25H, device selection affordance 2506A identifies the Living Room device as "Tim's Living Room" (e.g., wherein "Tim" is a user associated with device 2500).

[0735] In some embodiments, playback of media by guests (e.g., by device 2502) on a device of the home set of devices does not affect a preference profile and/or playback history of a user associated with the set of devices. For example, FIG. 25I illustrates an exemplary interface 2520 for browsing the recently played media items by the user Tim (e.g., the user of device 2504). In some embodiments, device 2504 displays interface 2520 subsequent to user input granting permission to a device that is not a member of the set of devices (e.g., to device 2502). As shown, the recently played media items 2520A-2520C do not include the media item "Same Thing" that was played back in response to the request by device 2502.

[0736] FIG. 26 is a flow diagram illustrating a method for allowing a guest device to initiate playback using an electronic device in accordance with some embodiments. Method 2600 is performed at a device (e.g., 100, 300, 500, 580). Some operations in method 2600 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0737] As described below, method 2600 provides an intuitive way for allowing a guest device to initiate playback. The method reduces the cognitive burden on a user for allowing a guest device to initiate playback, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to allowing a guest device to initiate playback faster and more efficiently conserves power and increases the time between battery charges.

[0738] The electronic device (e.g., 2500), which is a first device, receives a request (2602), from a third device (e.g., 2502), to initiate media playback on at least one device of a set of devices (e.g., a set that includes devices 2500 and 2504).

[0739] In some embodiments, the set of devices comprises (2604) the first device (e.g., 2500) and a second device (e.g., 2504). In some embodiments, the set of devices is a home set of devices (e.g., configured using a home control application as described with respect to FIG. 6A). For example, the set of devices includes one or more devices associated with one or more members of a household. In some embodiments, a member of the household is identified by device (e.g., from which a request originates) and/or by one or more user accounts associated with one or more devices (e.g., the user's devices are each logged into the user's iCloud account). Thus, a household can be defined as devices associated with a set of user accounts. In some embodiments, the set of user accounts comprises a single user account. In some embodiments, the set of user accounts comprises more than one user account.

[0740] In some embodiments, the third device is not a (2606) member of the set of devices. For example, the third device belongs to a guest or visitor to the home that includes the set of devices. For example, the third device can be associated with a user account that is not a member of the set of user accounts that comprise the set of users making up the household.), and

[0741] In some embodiments, the third device (e.g., 2502) is connected to (2608) the first device (e.g., 2500). In some examples, the devices are connected via Wi-Fi, Bluetooth, or the like.

[0742] In some embodiments, in response to receiving the request (2610): in accordance with a determination that the third device (e.g., 2502) has permission (e.g., previously granted by

member of the home) to initiate media playback on the at least one device (e.g., 2500) of set of devices, the electronic device (e.g., 2500) initiates (2612) media playback on the set of devices (e.g., 2500). In some embodiments, a set of devices includes one or more devices.

[0743] In accordance with a determination (2614) that the third device does not have permission to initiate media playback on the at least one device of set of devices: the electronic device (e.g., 2500) transmits (2616) a request for permission (e.g., represented by 2516 of FIG. 25F) to the second device (e.g., 2504). For example, the second device displays a prompt (e.g., as shown in 2516 of FIG. 25F).

[0744] Requesting permission, by a first device, from a second device to initiate playback based on a request from a third device in response to the request provides users with the ability to quickly and dynamically request permission for media playback while reducing the number of user inputs required to navigate permissions menus or resend a request to initiate playback after being granted permission. Providing improved media playback permissions features to the user enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0745] In response to receiving permission from the second device, the electronic device (e.g., 2500) initiates (2618) media playback on the at least one device (e.g., 2500) of the set of devices. In some embodiments, initiating playback includes transmitting an instruction (e.g., to another device in the set of devices) to initiate playback.

[0746] In some embodiments, in accordance with a determination that a first user account associated with the third device has permission to initiate media playback on the set of devices, the electronic device (e.g., 2500) determines that the third device (e.g., 2502) has permission (e.g., device 2502 has permission by virtue being associated with a user (e.g., user account) that has permission), wherein the first user account is not associated with any device of the set of devices (e.g., is an account (e.g., iCloud) of a guest user that is not associated with any device in the home of the home set of devices). In accordance with a determination that the first user account associated with the third device does not have permission to initiate media playback on

the set of devices, the electronic device (e.g., 2500) forgoes determining that the third device has permission. For example, if the user account associated with the third device does not have permission, playback does not initiate in response to a request from the third device.

[0747] In some embodiments, further in accordance with a determination that the third device does not have permission, the electronic device (e.g., 2500) transmits the request for permission to a plurality of devices (e.g., 2504 and another device (e.g., a personal device with a display), both associated with the home owner) that includes the second device (e.g., 2504). For example, the plurality of devices are devices that are associated with the first device. In some embodiments, the first device and the plurality of devices are each associated with one or more user accounts that are not associated with the third device. For example, the plurality of devices can be other devices associated with an owner (e.g., their user account) of the first device (e.g., logged into a common account belonging to the owner, or logged in to accounts associated with a predefined group of users such as a family of the owner), but the third device is associated with a guest (e.g., with the guest's user account) to the owner's home. Thus, the third device is not included in the group of users in the owner's household.

[0748] In some embodiments, initiating media playback on the at least one device of the set of devices comprises: initiating playback on the first device (e.g., 2500); and forgoing initiating playback on the second device (e.g., 2504). Thus, in some examples, the device that initiates playback is not the device that granted permission.

[0749] In some embodiments, subsequent to receiving permission from the second device (e.g., 2504): the electronic device (e.g., 2500) receives a request, from the third device (e.g., 2502) , to initiate playback on a fourth device (e.g., another smart speaker device that is not device 2500), wherein the set of devices comprises the fourth device, and wherein the fourth device is not included in the at least one device of the set of devices (e.g., the original request (by device 2502) to initiate playback on device 2500 did not include a request to initiate playback on the fourth device.). In response to receiving the request to initiate playback on a fourth device, (by device 2502) initiates playback on the fourth device. In some embodiments, initiating playback includes transmitting an instruction (e.g., to another device) to initiate playback. For example, if the guest user subsequently selects additional devices (e.g., at a multi-device

interface) in the home owner's home (e.g., after being granted permission to initiate playback on the first device 2500), the permission extends to other devices (e.g., all devices) associated with the home set of devices.

[0750] In some embodiments, subsequent to receiving permission from the second device: receiving a request, from the third device, to remove a media item (e.g., as shown in FIGS. 16M-16O) from a playback queue on the at least one device of the set of devices (e.g., 2500). In response to receiving the request: in accordance with a determination that the media item was added to the playback queue by the third device (e.g., was added by the guest), the electronic device (e.g., 2500) removes the media item from the playback queue. In accordance with a determination that that the media item was not added to the playback queue by the third device (e.g., not added by the guest), forgoing removing the media item from the playback queue.

[0751] In some embodiments, the first device is associated with a media preference profile (e.g., 2316A as shown in FIG. 23C), and further in response to receiving the request from the third device that is not a member of the set of devices, the electronic device (e.g., 2500) forgoes updating the media preference profile based on the request. For example, device 2500 (e.g., which is associated with the home owner) does not update a taste profile associated with the owner's user account based on media items played back in response to a request for another user (e.g., a user associated with a guest user account).

[0752] Note that details of the processes described above with respect to method 2400 (e.g., FIG. 24) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 2400. For brevity, these details are not repeated below.

[0753] FIGS. 27A-27E illustrate exemplary user interfaces for providing data based on a physical position, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 28.

[0754] FIG. 27A illustrates an exemplary scenario in which two devices (2700 and 2702) are connected. In some embodiments, device 2700 includes one or more features of device 100,

300, 500, and 580. In some embodiments, device 2702 includes one or more features of device 100, 300, 500, and 580. In some examples, a user may wish to use device 2700 (e.g., a smart speaker) as an extension of device 2702 (e.g., a personal device), wherein device 2702 is associated with the user's account (e.g., a thus their account data). For example, utilizing a smart speaker as an extension of a device (e.g., to output audio, phone calls, dictated messages, or the like; and/or to detect user voice input) allows the user improved hands-free access to the functions of their personal device. Further example, such functionality can also extend the distance that a user can be from their personal device while maintaining the ability to access data from their personal device. However, providing unconditional access of data from a personal device to a public device (e.g., potentially used by more than one user in a household) can be undesirable to the owner of the personal device. Thus, access to data from the personal device can be subject to satisfaction of one or more conditions, such as physical position condition. FIGS. 27B-27E provide exemplary interfaces for facilitating the sharing of data from one device to another based on physical positioning.

[0755] FIG. 27B illustrates an exemplary scenario in which the physical positioning of a device (e.g., 2702) satisfies a permissive condition (e.g., is placed in close proximity to device 2700). In the example of FIG. 27B, device 2700 received a user input request (e.g., a voice input), requesting the output of data from personal device 2702 through device 2700. For example, the request is a user input voice request to dictate the user's messages stored on (or otherwise associated with) device 2702. In response to the request, device 2700 determines that a physical positioning of device 2702 is in satisfaction of a physical position-based permissive condition. For example, device 2700 determines that device 2702 is in a permissive physical position (e.g., is in close proximity, such as within proximity zone 2704). In response to determining that the physical positioning of device 2702 satisfies a permissive condition, device 2700 outputs the data (e.g., dictation output 2706). Dictation output 2706 includes the response "You have 3 appointments today." In this example, dictation output 2706 is output in response to the user input voice request "Do I have any appointments today?"

[0756] FIG. 27C illustrates an exemplary scenario in which the physical positioning of a device (e.g., 2702) does not satisfy a permissive condition. As shown, for example, device 2702 is not within proximity zone 2704 of device 2700, which is the permissive condition in this

example. Thus, in response to receiving voice input "Do I have any appointments today?", device 2700 does not output the dictation shown in FIG. 27B. In some embodiments, a device (e.g., 2700) outputs an indication of how to satisfy the permissive condition. As shown, device 2700 outputs the dictation 2708 "Please move your phone closer".

[0757] As illustrated above, a permissive condition can be based on device proximity. In some embodiments, a permissive condition is based on (or further based on) an orientation of a device (e.g., 2702). FIG. 27D illustrates an exemplary scenario in which device 2702 is within close proximity to device 2700, but does not satisfy a permissive condition. In the example shown in FIG. 27D, the permissive condition is satisfied when device 2702 is in a permissive physical orientation. In some embodiments, an orientation of a device (e.g., 2702) satisfies the permissive condition when the device is face up. For example, in FIG. 27D, the main display (e.g., the "face" or "front" of the device) the main display 2703 of device 2702 is oriented down (e.g., the device is face down). In some embodiments, an orientation of a device satisfies the permissive condition when the device (e.g., 2702) is pointing in a direction (e.g., from the top of the device) toward the second device (e.g., 2700). In some embodiments, a permissive condition is satisfied when both a physical proximity and a physical orientation condition are satisfied.

[0758] FIG. 27E illustrates an exemplary scenario in which the physical positioning of device 2702 satisfies the permissive condition. FIG. 27E illustrates that device 2700 can be used to output dictation of data from device 2702. As shown, device 2700 outputs a dictation of a text message, the dictation being: "Text message from Bob, 'Let's meet at 6 P.M.'".

[0759] FIG. 28 is a flow diagram illustrating a method for providing data based on a physical position using an electronic device in accordance with some embodiments. Method 2800 is performed at a device (e.g., 100, 300, 500, 580). Some operations in method 2800 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0760] As described below, method 2800 provides an intuitive way for providing data based on a physical position. The method reduces the cognitive burden on a user for providing data based on a physical position, thereby creating a more efficient human-machine interface. For

battery-operated computing devices, enabling a user to provide data based on a physical position faster and more efficiently conserves power and increases the time between battery charges.

[0761] The electronic device (e.g., 2700), which is at a first device associated with a user account (e.g., logged into user's iCloud) receives (2802) user input (e.g., a user input voice command) representing a request to access data from the user account. In some embodiments, data from a user account can include one or more of: calendar data (e.g., meetings, events, or the like), messages (e.g., text messages, iMessages, emails, voice messages, or the like), notes, reminders (e.g., audio or text reminders), or the like.

[0762] In response to receiving (2804) the user input (e.g., user input voice command), the first device (e.g., 2700) determines (2806) whether a physical positioning of the first device (e.g., 2700) and a second device (e.g., 2702) associated with the user account satisfies a permissive condition. In some embodiments, a permissive condition is satisfied when the second device is in close proximity to the first device (e.g., as shown in FIG. 27B). In some embodiments, a permissive condition is satisfied when a device orientation of the second device is a permissive orientation (e.g., lying face up on a surface) (e.g., as shown in FIG. 27B).

[0763] In accordance with a determination that the physical positioning satisfies the permissive condition, the first device (e.g., 2700) provides (2808) the data from the user account through the first device (e.g., dictates the data, as shown in FIG. 27B).

[0764] In accordance with a determination that the physical positioning does not satisfy the permissive condition, the first device (e.g., 2700) forgoes (2810) providing the data from the user account through the first device (e.g., requests that the second device is placed in close proximity, as shown in FIG. 27C; requests unlocking of second device, as shown in FIG. 27D).

[0765] Using a determination of whether the physical positioning of device satisfies a permissive condition provides the user with an easy technique for granting permission without extraneous user inputs. Performing an operation when a set of conditions has been met without requiring further user input enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage

and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0766] In some embodiments, providing the data from the user account through the first device comprises: outputting an audio signal (e.g., 2706 or 2712) representative of the data from the user account.

[0767] In some embodiments, outputting an audio signal representative of the data from the user account comprises: causing audible output of a dictation (e.g., 2706 or 2712) representative of the data from the user account using a text-to-speech algorithm In some embodiments, the dictation is in a natural language (e.g., a spoken language such as English, Spanish, French, or the like). For example, the dictation comprises data being read aloud (e.g., using a text-to-speech algorithm) through a speaker of the first device.

[0768] In some embodiments, the data from the user account comprises data from one or more of: a calendar entry (e.g., a meeting, an appointment), a notification (e.g., of an incoming text message), an email, a text message, a stored text document (e.g., a note), stored contact information, phone call audio, video call audio or video (e.g., FaceTime, Skype), a voice message, an address (e.g., user's home), and a media item (e.g., music, video).

[0769] In some embodiments, determining whether the physical positioning of the first device and second device satisfies the permissive condition comprises: determining whether a physical proximity between the first device and the second device satisfies a proximity condition (e.g., the first and second device are in close proximity to each other).

[0770] In some embodiments, determining whether the physical positioning of the first device and second device satisfies the permissive condition comprises: determining whether a physical orientation of the first device satisfies an orientation condition. In some embodiments, the physical orientation can be detected via sensors on the second device (e.g., and communicated to the first device), such as one or more of: accelerometer, ambient light sensor, camera, or the like. In some embodiments, an orientation condition is satisfied when the second device is positioned such that a main display is not covered. For example, a device that is resting face up (e.g., with its display visible, in contrast to face down) satisfies the orientation condition.

[0771] In some embodiments, further in accordance with a determination that the physical positioning does not satisfy the permissive condition, the first device (e.g., 2700) outputs an indication (e.g., 2708 or 2710) that the permissive condition is not satisfied (e.g., the devices are not in close proximity).

[0772] In some embodiments, outputting the indication that the permissive condition is not satisfied comprises: outputting an audible indication (e.g., 2708 or 2710) of a user action required to satisfy the permissive condition (e.g., a dictated output of: "Please unlock your device" or "Please bring your phone closer", or the like.).

[0773] In some embodiments, the user input representing a request to access data from the user account is a first user input, the method further comprising, subsequent to providing the data from the user account through the first device (e.g., as shown in FIG. 27B): receiving a second user input (e.g., a user input voice command) representing a request to access additional data from the user account. In response to receiving the second user input, determining whether the physical positioning of the first device (e.g., 2700) and the second device (e.g., 2702) associated with the user account satisfies the permissive condition. In accordance with a determination that the physical positioning satisfies the permissive condition, providing the additional data (e.g., as shown in FIG. 27E) from the user account through the first device. In accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the additional data from the user account through the first device (e.g., as shown in FIG. 27C).

[0774] Note that details of the processes described above with respect to method 2800 (e.g., FIG. 28) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 2800. For brevity, these details are not repeated below.

[0775] FIGS. 29A-29I illustrate exemplary user interfaces and techniques for manage playback of media on a plurality of devices, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 30.

[0776] FIG. 29A illustrates an exemplary multi-device interface (e.g., such as multi-device interface 610 or 810 as described above with respect to FIGS. 6 and 8) that includes Devices 1 through 4, displayed on display 2902 of device 2900. In this example, a user uses device 2900 to transmit a request to initiate media playback on a set of devices that includes Devices 1 through 4. In some embodiments, device 2900 includes one or more features of devices 100, 300, 500, and 580.

[0777] FIG. 29B illustrates an exemplary arrangement of Devices 1 through 4 (e.g., devices 2910, 2912, 2914, and 2916, respectively) that are included in the set of devices on which playback has been requested (e.g., by device 2900). In some embodiments, devices 2910, 2912, 2914, and 2916 are smart speakers that include one or more features of devices 100, 300, 500, and 580.

[0778] At FIG. 29C, the set of devices have initiated playback of media (e.g., are outputting audio), in response to the request form device 2900.

[0779] In some embodiments, a set of devices that is concurrently playing back media (e.g., as part the same media session) synchronizes playback of the media. In some embodiments, one device coordinates playback synchronization (e.g., also referred to as a "primary device" or "hero device"). Coordinating playback synchronization can include one or more of: processing playback control commands (e.g., pause, next track), streaming media to the other devices in the set of devices, storing and/or maintaining a playback queue, detecting whether the other devices are outputting the media in synchronization with each other (e.g., via use of a clock signal), or the like.

[0780] In some embodiments, a set of devices (e.g., for a media session) includes an heir device (also referred to as a "designated heir" or a "backup primary device"). In some embodiments, the designated heir is configured to take over responsibility of the primary device in response to the primary device no longer being able to playback media (e.g., for the media session) and/or coordinate playback synchronization. In some embodiments, several "generations" of heirs are designated (e.g., a designated heir, a designated heir to the designated heir, and so forth).

[0781] FIG. 29D depicts an exemplary scenario in which device 2910 is the designated heir to device 2912, the primary device. As shown, device 2912 (primary) connects to each of the devices in the set and, for example, synchronizes playback from each device.

[0782] FIG. 29E depicts an exemplary scenario in which device 2910 has become the primary device and (e.g., because device 2912 is no longer available) and device 2914 is now the designated heir to device 2910, the primary device. For example, the scenario in FIG. 29E follows an event (e.g., during the scenario in FIG. 29D) in which device 2912 (e.g., the previous primary) is no longer available for playback during the media session on the set of devices. As shown, device 2910 (primary) connects to each of the remaining devices in the set (e.g., 2914 and 1216) and, for example, coordinates playback with each of these devices.

[0783] FIG. 29F depicts an exemplary scenario in which device 2914 has become the primary device and (e.g., because device 2910 is no longer available) and device 2916 is now the designated heir to device 2914, the primary device. For example, the scenario in FIG. 29F follows an event (e.g., during the scenario in FIG. 29E) in which device 2910 (e.g., the previous primary) is no longer available for playback during the media session on the set of devices). As shown, device 2914 (primary) connects to each of the remaining devices (e.g., 2916) in the set and, for example, coordinates playback with each of these devices.

[0784] FIG. 29G depicts an exemplary scenario similar to that shown in FIG. 29D, but in which device 2916 is not part of the original set of devices selected by the user (e.g., using device 2900) for the media playback session. In some embodiments, a device that is not part of the set of devices (e.g., of the requested media session) performs the function of designating the primary device and/or the designated heir for the media playback session. For example, as shown in FIG. 29G, device 2916, even though not in the set of devices, designates device 2910 as the heir and device 2912 as the primary device. In some embodiments, the device that designates a primary device is a home media control device (e.g., an Apple TV).

[0785] In some embodiments, primary device responsibility (e.g., playback coordination) includes processing playback control commands. For example, because the devices in a set of devices are interconnected with each other (and potentially connected with other devices), a playback control command can be received by any device of the set of devices (or by the other

devices). In some embodiments, the playback control command is forwarded to the primary device if it is a device that is not the primary device for a media session. In some embodiments, the primary device processes the playback control command, and then transmits instructions to the (e.g., remaining) devices in the set of devices (e.g., for enacting the playback control command).

[0786] For example, FIG. 29H illustrates the exemplary scenario in which a playback control command 2920 is received at a non-primary device (e.g., designated heir, device 2910 in this example). Thus, as shown, device 2910 forwards the playback control command to the primary device. In contrast, at FIG. 29I, device 2910 is the primary device and receives the playback control command 2920. Thus, in the example depict in FIG. 29I, device 2910 processes the playback control command and transmits playback instructions (e.g., skip to next track, pause playback, or the like) to the remaining devices in the set (e.g., that are included in the media session).

[0787] FIG. 30 is a flow diagram illustrating a method for managing playback coordination using an electronic device in accordance with some embodiments. Method 3000 is performed at a device (e.g., 100, 300, 500, 580). Some operations in method 3000 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0788] As described below, method 3000 provides an intuitive way for managing playback coordination. The method reduces the cognitive burden on a user for managing playback coordination, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to manage playback coordination faster and more efficiently conserves power and increases the time between battery charges.

[0789] The electronic device (e.g., 2910), which is a first device, receives (3002) a command to initiate playback of media on a set of devices comprising the first device (e.g., 2910) and a second device (e.g., 2912). In some embodiments, the first and second devices are both configured to be capable of receiving playback coordination responsibility for the set of devices.

[0790] In response to receiving the command, the first device (e.g., 2910) initiates (3004) synchronized playback of the media on the set of devices (e.g., 2910 and 2912), wherein the second device is designated (3006) as a primary device (e.g., as shown in FIG. 29D) configured to coordinate the playback of the media on the set of devices, and wherein the first device is configured (3008) to be a designated heir to the primary device (e.g., as shown in FIG. 29D). In some embodiments, coordinating playback on the set of devices includes one or more of: storing a media playback queue, synchronizing media playback start time, and receiving and routing playback commands. In some embodiments, a designated heir to the primary device is a device that will take over duty as the primary device if the current primary device becomes unavailable.

[0791] Initiating synchronized playback and automatically managing the coordination of media playback without requiring further user input reduces the burden on the user when playing back media on a plurality of devices. Performing an operation when a set of conditions has been met without requiring further user input enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0792] In some embodiments, the set of devices further comprises a third device (e.g., 2914 as shown in FIG. 29D), and during playback on the set of devices, the first device (e.g., 2910) detects that the second device (e.g., 2912) is no longer available for playback of the media (e.g., as shown in FIG. 29E). For example, a device is no longer available for playback if the device is requested to play something else or to cease playback, the device is disconnected, or the like. In response to detecting that the second device is no longer available to playback the media: the first device (e.g., 2910) determines that the first device is designated as the primary device configured to coordinate the playback of the media on the set of devices (e.g., as shown in FIG. 29E), continues playback of the media on the first device (e.g., as shown in FIG. 29E), and coordinates the playback of the media on the third device (e.g., 2914, as shown in FIG. 29E and 29I). In some embodiments, the first device determines that it is the primary device. In some embodiments, the first device receives notification from another device that the first device has been designated the primary device. For example, the first device assumes the role of primary

device and begins coordination playback on the remaining device(s) in the set of devices. In this example, the third device is a remaining device.

[0793] In some embodiments, coordinating the playback of the media on the third device comprises one or more of: providing a link to the media (e.g., a URL); providing a stream of the media (e.g., a real time or buffered stream of data representing the media); and providing a copy of the media (e.g., a media file).

[0794] In some embodiments, the first device (e.g., 2910) receives a command to cease playback of media on the first device. In response to receiving the command to cease playback of media on the first device: the first device (e.g., 2910) ceases playback of media on the first device (e.g., as shown in FIG. 29F), and designates the third device to be the primary device configured to coordinate the playback of the media on the set of devices (e.g., as shown in FIG. 29F).

[0795] In some embodiments, the set of devices further comprises a third device (e.g., 2914), and during playback on the set of devices (e.g., 2910, 2912, and 2914), the first device (e.g., 2910) detects that the second device (e.g., 2912) is no longer available for playback of the media (e.g., as shown in FIG. 29E). In response to detecting that the second device is no longer available to playback the media: the first device (e.g., 2910) determines that the third device (e.g., 2914) (e.g., despite the first device being the designated heir) is designated as the primary device configured to coordinate the playback of the media on the set of devices, and the first device (e.g., 2910) continues playback of the media on the first device, and the first device (e.g., 2910) receives, from the third device (e.g., the new primary device), an instruction for coordinating playback of the media on the first device. For example, the first device does not assume the role of primary device, and thus begins receiving coordinating instructions for playback from the new primary device, the third device.

[0796] In some embodiments, the first device (e.g., 2910) receives, from a fourth device (e.g., 2916), a command to initiate playback of media on the set of devices. The first device (e.g., 2910) receives, from the fourth device, a designation that the first device is the designated heir. For example, the device that transmits the command to initiate playback and that

designates the designated heir is a device different than the designated heir and the designated primary device.

[0797] In some embodiments, the fourth device is not a member of the set of devices. For example, the fourth device is a device in the home set of devices that initially coordinates playback on device(s) regardless of whether playback includes the fourth device. In some embodiments, initially coordinating playback includes designating a designated heir and/or a primary device for the playback session on the set of devices.

[0798] In some embodiments, initiating synchronized playback of the media on the set of devices comprises transmitting a copy of a playback queue to each device in the set of devices.

[0799] In some embodiments, during synchronized playback, the first device (e.g., 2910) receives, at the first device, a playback control command (e.g., 2920, such as pause, next track, seek, or the like). In response to receiving the playback control command: in accordance with the first device being designated the primary device, the first device (e.g., 2910) transmits an instruction to each of the devices in the set of devices based on the playback control command (e.g., as shown in FIG. 29I). In accordance with the first device not being designated the primary device, the first device (e.g., 2910) transmits the playback control command to a designated primary device (e.g., as shown in FIG. 29H). For example the first device forwards the command to the primary device for processing, and does not take further action until receiving an instruction from the primary device).

[0800] Note that details of the processes described above with respect to method 3000 (e.g., FIG. 30) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 3000. For brevity, these details are not repeated below.

[0801] FIGS. 31A-31N illustrate exemplary user interfaces for outputting an audible output based on user location, in accordance with some embodiments. The user interfaces in these figures are used to illustrate the processes described below, including the processes in FIG. 32.

[0802] FIG. 31A illustrates in exemplary home that includes an exemplary home set of devices (e.g., the set of devices 3100, 3102, and 3104). As shown, devices 3100 and 3102 are in the room named "Living Room" (e.g., are named or otherwise associated with an identifier "Living Room") and device 3104 is in the room "Bedroom". In this example, devices 3100, 3102, and 3104 represent smart speaker devices that the user has positioned around their home, and that are connected to each other (e.g., via Wi-Fi). In some embodiments, devices 3100, 3102, and 3104 include one or more features of devices 100, 300, 500, and 580. FIG. 31A also depicts device 3106 (representing a car), which is optionally a device that is connected to the user's home set of devices (e.g., devices 3100, 3102, and 3104). In some embodiments, device 3108 includes one or more features of devices 100, 300, 500, and 580.

[0803] As described above, devices such as devices 3100, 3102, and 3104 (e.g., smart speaker devices), in accordance with some embodiments, can determine a spatial position of a user for targeting audio content to a voice command source and/or act as an extension (e.g., for notifications output) of another device (e.g., a personal device). Similarly, a device (or a set of devices) (e.g., such as devices 3100, 3102, and 3104) can be configured to automatically and selectively output audible output based on whether a user is within a range of the device (e.g., physically present near the device). FIGS 31B-31N illustrates exemplary techniques for doing so.

[0804] FIG. 31B depicts the scenario of FIG. 31A, however user 3108 is present in the Living Room near device 3100. In some embodiments, a device (e.g., 3100) of a set of devices (e.g., 3100, 3102, and 3104) receives an audible output trigger event. For example, in FIG. 31B, user 3108 utters the user input voice command 3112 "Hey Siri, set a timer for 10 minutes". In some embodiments, an audible output trigger event is any data configured to cause a device to output an audible output in response to a trigger event. For instance, in some examples, an audible output trigger event includes one or more of: notification settings (e.g., identifying events for which a device should issue an audible notification such as phone calls, text messages, reminders, calendar events, or the like), a user input request to set an alarm/reminder/timer, or the like. In some embodiments, the audible output trigger event is received via user input (e.g., a voice command) or data transfer (e.g., from a user's personal device, or from their user account (e.g., iCloud)). Subsequent to the user input voice command 3112, in the example shown in FIG.

31C, the user is still in the same position nearest device 3100, and so device 3100 outputs an audible alarm 3110 upon the occurrence of the trigger event (e.g., expiration of the 10 minute timer).

[0805] In some embodiments, a device (e.g., 3100) has an audible output range (e.g., a designated, predetermined, or dynamic range). In some embodiments, if a device (e.g., 3100) determines that a user is located outside of the audible output range of the device, it forgoes outputting the audible output when the trigger event occurs. For example, as shown in FIG. 31D, user 3108 is closest to device 3102 when the trigger event occurs (e.g., the expiration of the 10 minute timer), and thus device 3102 outputs the audible alarm 3110. Similarly, in FIG. 31E, the user 3108 is closest to device 3104 when the trigger event occurs (e.g., the expiration of the 10 minute timer), and thus device 3104 outputs the audible alarm 3110.

[0806] In some embodiments, determining whether a user is located within an audible output range is based on one or more of: a detected physical proximity of the device (e.g., 3100) to the user (e.g., 3108), a detected presence of the user (e.g., 3108) in the same room as the device (e.g., 3100), and the output characteristics (e.g., audio output power) of the device (e.g., 3100). In some embodiments, determining whether a user is located within an audible output range is based on detection of a device associated with the user.

[0807] An audible output range can be determined in a variety of manners. In some embodiments, the audible output range (e.g., of each device) is fixed. For example, FIG. 31F depicts the audible output range 3100A of device 3100 (e.g., is a proximity range represented by the dotted line), the audible output range 3102A of device 3102, and audible output range 3104A of device 3104, which represent fixed ranges. In some embodiments, an audible output range is independent for each device (or is not affected by whether there are other devices in the set of devices, or other devices nearby. For example, a device can have a fixed proximity range (e.g., 3100A, 3102A, or 3104A), such that it is possible for a user to be out of range (e.g., in the garage in FIG. 31F) of each device in a set of devices (e.g., 3100, 3102, and 3104).

[0808] In some embodiments, a user can be within the audible output range of a plurality of devices. For example, as shown in FIG. 31G, user 3108 is within the output range of both devices 3100 and 3102, and thus both devices are outputting audible alarm 3110. In some

embodiments, only one device of the plurality of devices outputs an audible output if a user is within range of the plurality of devices.

[0809] In some embodiments, an audible output range of a device is the room in which a device is located. For example, as shown in FIG. 31H, the user is physically positioned in the Bedroom upon the occurrence of the trigger event, and thus device 3104 outputs audible alarm. Notably, in FIG. 31H, device 3104 outputs the alarm even though user 3106 is physically positioned closer to device 3100. Thus, in this example, it is the room that the user is in, and not necessarily physical proximity, that affects whether a device outputs in response to a trigger event.

[0810] In some embodiments, each device in a set of devices (e.g., 3100, 3102, and 3104) receives the audible output trigger event. As shown in FIG. 31H, even though the user's voice input (e.g., to set a timer) was received at device 3100, device 3104 responds to the trigger event by audibly outputting the alarm 3110. In some embodiments, a device (e.g., 3104) receives an audible output trigger event from another device (e.g., another device (e.g., 3100) in a set of devices, such as a home set of devices; or another device connected to the device). In some embodiments, a device (e.g., 3104) receives an audible output trigger event from a user account (e.g., downloaded from an iCloud account).

[0811] In some embodiments, the audible output range of a device (e.g., 3100) is relative to one or more other devices (e.g., 3102), or is otherwise affected by other devices in a set of devices or nearby. For instance, FIG. 31I illustrates a scenario in which device 3100 is physically nearest to the user 3106 (or is the last device to which the user 3106 was detected to be nearest). Thus, in FIG. 31I, device 3100 outputs the audible alarm 3110, even though the user is in the next room. This, for example, can ensure that at least one device outputs the audible output. In some embodiments, a device adjusts its output (e.g., outputs at a loud volume) based on a user's positioning (e.g., if the user is detect to be positioned far away from the device).

[0812] In some embodiments, a device (e.g., 3100) of the set of devices (e.g., 3100, 3102, and 3104) detects a user voice to determine whether the user is located within the device's audible output range. For example, in FIG. 31J user 3108 speaking with user 3116 in the Living Room, as represented by user voice 3114 ("How are you?"). In some example, device 3100

detects the voice of user 3106, and determines that the user is in the Living Room (e.g., and within the audible output range of device 3100). Thus, in this example, device 3100 will output the audible alarm 3110 upon occurrence of the trigger (e.g., if user 3106 continues to be detected, or is detected within a threshold amount of time preceding the trigger event (e.g., within the last 5 minutes before the trigger)).

[0813] In some embodiments, a device (e.g., 3106) that is not one of the set of devices (e.g., 3100, 3102, and 3104) is connected to the set of devices. In this example, device 3106 is a car that connected to a personal device of user 3106 when the user entered the car. In some embodiments, the devices of the set of devices forgo outputting audible outputs in response to determining that the user is within an audible output range (e.g., while the car is powered on) of the device (e.g., 3106) that is not one of the set of devices. For example, as shown in FIG. 31K, when the user enters car 3106, the set of devices 3100, 3102, and 3104 do not output the audio output in response to trigger event (e.g., because the user is likely to be leaving their home).

[0814] In some examples, the devices of a set of devices are used as intelligent end-to-end communication tools. In some embodiments, a first device (e.g., 3100) of a set of devices (e.g., 3100, 3102, and 3104) receives a user input request (e.g., 3118) to deliver a message (e.g., 3120) to a user (e.g., 3116) that is within audible output range of at least one device of the set of devices. In some embodiments, the first device (e.g., 3100) transmits the message to a second device (e.g., 3104) (e.g., that the user 3116 is within audible output range of) of the set of devices (e.g., 3100, 3102, and 3104). For example, as shown in FIG. 31L, user 3108 is located in the Living Room and user 3116 is located in the Bedroom. At FIG. 31M, user 3108 utters the user input voice command 3118 "Send a message to Tim, dinner is ready", which is detected by device 3100. As described above, the devices of the set of devices (e.g., 3100, 3102, and 3104) can determine a physical position of a user in order to deliver output of audible notifications as appropriate. Thus, as shown in FIG. 31N, a message 3120 ("Dinner is ready") is audibly output by a second device (e.g., 3104) of the set. In this example, the user 3116 (the target of the voice input command, Tim) is located within the audible output range of device 3104, and thus device 3104 outputs the message 3120. In some embodiments, the first device transmits a request to the second device inquiring whether the user (e.g., 3116) is within the devices audible output range (e.g., and if confirmation is received, transmits the message 3120 to the second device).

[0815] In some embodiments, the audible output range of a device (e.g., 3100) is based on one or more of: other detectable indicia of a user presence or location, including footsteps (e.g., determining and detecting a footstep pattern associated with a user, tracking the location of a set of footsteps as they move around (e.g., which are presumed to be the user), determining and detecting a breathing pattern of the user, using a visual detection mechanism (e.g., a motion sensor, a camera), or the like.

[0816] FIG. 32 is a flow diagram illustrating a method for outputting an audible output based on user location using an electronic device in accordance with some embodiments. Method 3200 is performed at a device (e.g., 100, 300, 500). Some operations in method 3200 are, optionally, combined, the order of some operations are, optionally, changed, and some operations are, optionally, omitted.

[0817] As described below, method 3200 provides an intuitive way for outputting an audible output based on user location. The method reduces the cognitive burden on a user for outputting an audible output based on user location, thereby creating a more efficient human-machine interface. For battery-operated computing devices, enabling a user to outputting an audible output based on user location faster and more efficiently conserves power and increases the time between battery charges.

[0818] The electronic device (e.g., 3100), which is a first device of a set of devices (e.g., 3100, 3102, and 3104) associated with a first user account: receives (3202) an audible output trigger event. In some embodiments, an audible output trigger event is any data configured to cause a device to output an audible output in response to a trigger event. In some embodiments, an audible output trigger events is one or more of: a notification (e.g., alarm/timer/reminder) for a particular time (e.g., 5:00PM) or some length of time in the future (e.g., in 30 minutes). In some embodiments, an audible output trigger event is a device notifications setting.

[0819] The first device (e.g., 3100) determines (3204) that a first user (e.g., 3108), associated with the first user account, is located within an audible output range (e.g., 3100A, 3102A, and/or 3104A) of a device of the set of devices. In some embodiments, an audible output range is independent for each device (or is not affected by whether there are other devices in the set of devices). For example, each device has an “fixed” proximity range, so a user (e.g., 3108) can be

out of range of each device in a set of devices. In some embodiments, the audible output range of a device is relative or otherwise affected by other devices in the set of devices. For example, the user may always be within the audible output range of the closest (most proximate) device of the set (e.g., closest device responds, even if it is in the next room or 100 meters away; a shown in FIG. 31I). In some embodiments, the audible output range is the room that the device is located in (e.g., for a device located in a room that is correctly identified as being in the "Bedroom" in device settings, then a user is within the audible output range of the bedroom device (e.g., 3104) is present (physically located) in the Bedroom, as shown in FIG. 31H.);

[0820] In accordance with a determination that the first user (e.g., 3108) is within the audible output range of the first device (e.g., 3100), and in accordance with an occurrence of the audible output trigger event, the first device (e.g., 3100) outputs (3206) an audible output (e.g., 3110 as shown in FIG. 31C). For example, the first device outputs an audible tone for timer expiring, or outputs a received phone call.

[0821] Using a determination of whether the physical position of a user is within an audible output range of a device provides the user with an easy technique for receiving output from appropriately positioned devices without extraneous user inputs. Performing an operation when a set of conditions has been met without requiring further user input enhances the operability of the device and makes the user-device interface more efficient (e.g., by helping the user to provide proper inputs and reducing user mistakes when operating/interacting with the device) which, additionally, reduces power usage and improves battery life of the device by enabling the user to use the device more quickly and efficiently.

[0822] In accordance with a determination that the first user is within the audible output range of a second device (e.g., 3102 as shown in FIG. 31D, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output (e.g., 3100 does not output audible tone for timer expiring; as shown in FIG. 31D).

[0823] In some embodiments, the device (e.g., 3102) of the set of devices is a second device, and wherein outputting the audible output comprises transmitting an instruction to the second device (e.g., 3102) to output the audible output (e.g., 3110 as shown in FIG. 31D).

[0824] In some embodiments, outputting the audible output includes playing an audible tone (e.g., a sound, or voice dictation). In some embodiments, the device also outputs a visual output (e.g., using an LED or a display.).

[0825] In some embodiments, receiving the audible output trigger event comprises receiving user input representing configuration of the audible output trigger event (e.g., setting a timer, an alarm, or reminder; or setting up notification or phone call output settings).

[0826] In some embodiments, receiving the audible output trigger event comprises receiving data representing audible output settings associated with the first user account (e.g., whether to push one or more of: phone calls, text message notifications, email notifications, etc., to the first device (e.g., for audible output)).

[0827] In some embodiments, the audible output range of the first device is based on one or more of: a detected physical proximity (e.g., distance) of the first device to the first user; a detected presence of the first user in the same room as the first device; and the output characteristics of the first device (e.g., speaker output power).

[0828] In some embodiments, determining that the first user is located within the audible output range of the device of the set of devices comprises one or more of the following: detecting, at the device (e.g., 3100), a voice associated with the first user (e.g., 3112 or 3114) (e.g., within a threshold time before the notification trigger event, indirectly (e.g., overhearing user), or directly (e.g., user interacts with device with a voice command)); detecting, at the device, a third device (e.g., a personal device) associated with the user, wherein the third device is not a member of the set of devices (e.g., the device is the user's personal phone, and is not part of a group of stationary devices (e.g., devices 3100, 3102, and 3104)); and receiving a signal, from an external sensor (e.g., a sensor on an external device, or a sensor on the device), representative of the first user being physically near the device (e.g., motion sensor, camera).

[0829] In some embodiments, subsequent to outputting the audible output: in accordance with a determination that the first user (e.g., 3108) is within an audible output range of a fourth device (e.g., 3106) that is not a member of the set of devices (e.g., 3100, 3102, and 3104), and in accordance with an occurrence of the audible output trigger event, the first device (e.g., 3100)

forgoes outputting the audible output. For example, the user has entered their car in the garage (e.g., which is connected to the set of devices), and the set of devices forgoes outputting the audible output even though the user is still be within range of a device of the set of devices.

[0830] In some embodiments, the first device (e.g., 3100) plays back media content. Further in accordance with a determination that the first user is within the audible output range of the fourth device (e.g., 3106) that is not a member of the set of devices, ceasing playback of the media content.

[0831] Note that details of the processes described above with respect to method 3200 (e.g., FIG. 32) are also applicable in an analogous manner to the methods described below/above. For example, method 900 optionally includes one or more of the characteristics of the various methods described above with reference to method 3200. For brevity, these details are not repeated below.

[0832] The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the techniques and their practical applications. Others skilled in the art are thereby enabled to best utilize the techniques and various embodiments with various modifications as are suited to the particular use contemplated.

[0833] Although the disclosure and examples have been fully described with reference to the accompanying drawings, 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 claims.

[0834] As described above, one aspect of the present technology is the gathering and use of data available from various sources to improve the delivery to users of invitational content or any other content that may be of interest to them. The present disclosure contemplates that in some instances, this gathered data may include personal information data that uniquely identifies or can be used to contact or locate a specific person. Such personal information data can include

demographic data, location-based data, telephone numbers, email addresses, home addresses, or any other identifying information.

[0835] The present disclosure recognizes that the use of such personal information data, in the present technology, can be used to the benefit of users. For example, the personal information data can be used to deliver targeted content that is of greater interest to the user. Accordingly, use of such personal information data enables calculated control of the delivered content. Further, other uses for personal information data that benefit the user are also contemplated by the present disclosure.

[0836] The present disclosure further contemplates that the entities responsible for the collection, analysis, disclosure, transfer, storage, or other use of such personal information data will comply with well-established privacy policies and/or privacy practices. In particular, such entities should implement and consistently use privacy policies and practices that are generally recognized as meeting or exceeding industry or governmental requirements for maintaining personal information data private and secure. For example, personal information from users should be collected for legitimate and reasonable uses of the entity and not shared or sold outside of those legitimate uses. Further, such collection should occur only after receiving the informed consent of the users. Additionally, such entities would take any needed steps for safeguarding and securing access to such personal information data and ensuring that others with access to the personal information data adhere to their privacy policies and procedures. Further, such entities can subject themselves to evaluation by third parties to certify their adherence to widely accepted privacy policies and practices.

[0837] Despite the foregoing, the present disclosure also contemplates embodiments in which users selectively block the use of, or access to, personal information data. That is, the present disclosure contemplates that hardware and/or software elements can be provided to prevent or block access to such personal information data. For example, in the case of advertisement delivery services, the present technology can be configured to allow users to select to "opt in" or "opt out" of participation in the collection of personal information data during registration for services. In another example, users can select not to provide location information

for targeted content delivery services. In yet another example, users can select to not provide precise location information, but permit the transfer of location zone information.

[0838] Therefore, although the present disclosure broadly covers use of personal information data to implement one or more various disclosed embodiments, the present disclosure also contemplates that the various embodiments can also be implemented without the need for accessing such personal information data. That is, the various embodiments of the present technology are not rendered inoperable due to the lack of all or a portion of such personal information data. For example, content can be selected and delivered to users by inferring preferences based on non-personal information data or a bare minimum amount of personal information, such as the content being requested by the device associated with a user, other non-personal information available to the content delivery services, or publically available information.

CLAIMS

What is claimed is:

1. A computer-implemented method, comprising:

at a device with a display:

receiving a user input; and

in response to receiving the user input, displaying, on the display, a multi-device interface that includes:

one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and

a media playback status of the plurality of available playback devices.

2. The method of claim 1, wherein the user input is a first user input, the method further comprising:

receiving a second user input representing selection of the one or more indicators associated with the plurality of available playback devices;

in response to receiving the second user input, transmitting an instruction to initiate playback on the plurality of available playback devices.

3. The method of any of claims 1-2, wherein the user input is a first user input, the method further comprising:

while displaying the multi-device interface, displaying, on the display, a first indicator of the one or more indicators;

receiving a third user input representing selection of the first indicator;

in response to receiving the third user input, displaying, on the display, affordances for each of the devices of the plurality of available playback devices.

4. The method of claim 3, further comprising, while the device is playing back a first media item:

receiving a fourth user input representing selection of a second indicator of the one or more indicators,

wherein the second indicator is associated with a set of playback devices that does not include the device;

subsequent to receiving the fourth user input representing selection of the second indicator, displaying, on the display, a browsing interface, wherein the browsing interface includes an affordance associated with a second media item;

receiving a fifth user input representing selection of the affordance associated with the second media item; and

in response to receiving the fifth user input:

transmitting an instruction to initiate playback of the second media item on the set of playback devices that does not include the device; and

continuing playing back the first media item on the device.

5. The method of claim 4, wherein the browsing interface includes an affordance associated with a third media item, the method further comprising:

receiving a sixth user input associated with the affordance associated with the third media item;

in accordance with receiving the sixth user input, displaying, on the display, the multi-device interface.

6. The method of any of claims 1-5, wherein displaying the multi-device interface comprises:

concurrently displaying, on the display:

a first region that includes the multi-device interface; and

a second region that includes content displayed prior to displaying the multi-device interface.

7. The method of claim 6, wherein the first region is visually overlaid on the second region.

8. The method of claim 6, further comprising:

receiving selection of an indicator of the one or more indicators, wherein the indicator includes a representation of a fourth media item; and

in response to receiving selection of the indicator, and while continuing to concurrently display the first and second regions, updating display of the second region to include at least a portion of the representation of the fourth media item.

9. The method of any claims 1-8, wherein the user input is a first user input, the method further comprising:

prior to receiving the user input:

receiving a seventh user input;

in response to receiving the seventh user input, displaying, on the display, a playback control interface that includes:

one or more playback control affordances;

a playback device selection affordance; and

an affordance for controlling a data connection of the device; and

receiving the first user input, wherein the first user input is associated with the playback control interface.

10. The method of claim 9, wherein the user input is a selection of the playback device selection affordance.

11. The method of claim 9, further comprising:

receiving the first user input, wherein the first user input has an input characteristic;

determining whether the input characteristic meets an input characteristic-dependent criteria;

in accordance with a determination that the input characteristic meets the input characteristic-dependent criteria, displaying, on the display, the multi-device interface; and

in accordance with a determination that the input characteristic intensity does not meet the input characteristic-dependent criteria, forgoing displaying, on the display, the multi-device interface.

12. The method of claim 9, wherein the seventh user input is received while displaying a user interface of an application executing on the device, the method further comprising:

while continuing to display at least a portion of the user interface of the application, displaying, on the display, the playback control interface; and

subsequent to receiving the first user input, and while continuing to display the at least a portion of the user interface of the application, displaying, on the display, the multi-device interface.

13. The method of any of claims 1-12, wherein the multi-device interface includes a volume control affordance associated with the plurality of available playback devices, the method further comprising:

receiving an eighth user input associated with the volume control affordance; and

in response to receiving the eighth user input, transmitting an instruction to adjust a playback volume of the plurality of available playback devices.

14. The method of any of claims 1-13, wherein the multi-device interface includes a plurality of individual volume control affordances, each associated with a single device of the plurality of available playback devices, the method further comprising:

receiving a ninth user input associated with an individual volume control affordance; and

in response to receiving the ninth user input, transmitting an instruction to adjust a playback volume of the associated device, wherein the associated device is different than the device.

15. The method of any of claims 1-14,

wherein the one or more indicators associated with the plurality of available playback devices includes:

a first indicator representing a fifth media item currently playing on the device, wherein the first indicator is displayed in a first region;

a second indicator representing a group of devices that are each currently playing back a sixth media item;

individual indicators for each device of the plurality of available playback devices; and

while displaying the multi-device interface, receiving a tenth user input representing a request to replace display of the first indicator in the first region with display of the second indicator; and

in response to the tenth user input, replacing display of the first indicator in the first region with display of the second indicator.

16. The method of claim 15, further comprising displaying an indication of whether each device of the plurality of available playback devices is currently playing back a media item associated with an indicator displayed in the first region.

17. The method of claim 15, further comprising:

receiving an eleventh user input associated with an individual indicator for a device of the plurality of available playback devices;

determining whether the eleventh user input represents a selection that exceeds a threshold duration;

in accordance with the eleventh user input representing a selection that exceeds the threshold duration, entering a volume adjustment mode for the device associated with the individual indicator; and

in accordance with the eleventh user input not representing a selection that exceeds the threshold duration, forgoing entering the volume adjustment mode for the device associated with the individual indicator.

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

receiving a twelfth user input representing a request to initiate media playback on the plurality of available devices;

in response to receiving the twelfth user input:

in accordance with a determination that the plurality of available devices includes at least one device configured to receive playback coordination responsibility, transferring playback coordination responsibility to the at least one device of the plurality of devices; and

in accordance with a determination that the plurality of available devices does not include at least one device configured to receive playback coordination responsibility, coordinating the media playback on the plurality of devices.

19. The method of claim 18, further comprising, while the device is coordinating the media playback on the plurality of devices:

receiving a thirteenth user input representing a request to initiate a phone call on the device;

in response to receiving the thirteenth user input representing the request to initiate the phone call:

initiating the phone call; and

continuing coordinating the media playback on the plurality of devices during the phone call.

20. The method of claim 18, further comprising, further in response to receiving the twelfth user input, transmitting data representing a playback queue to each of the devices in the plurality of devices.

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

receiving a fourteenth user input representing a request to initiate media playback on the plurality of available devices, wherein the fourteenth user input is associated with a first user;

in response to receiving the fourteenth user input:

in accordance with a determination that the first user is subject to a playback restriction, preventing the device from initiating media playback on the plurality of available devices in response to the eleventh user input; and

in accordance with a determination that the first user is not subject to the playback restriction, allowing the device to initiate media playback on the plurality of available devices in response to the eleventh user input.

22. The method of any of claims 1-21, further comprising:

receiving a fifteenth user input representing a request for media output on the device;

in response to receiving the fifteenth user input:

in accordance with a determination that the request for media output is a request to display content, transmitting an instruction to a preferred device of the plurality of available playback devices to initiate display of the content in response to the twelfth request; and

in accordance with a determination that the request for media output is not a request to display content:

forgoing transmitting the instruction to the preferred device of the plurality of available playback devices to initiate display of the content in response to the twelfth request; and

initiating media output on the device in response to the twelfth user input.

23. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for:

receiving a user input; and

in response to receiving the user input, displaying, on the display, a multi-device interface that includes:

one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and

a media playback status of the plurality of available playback devices.

24. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving a user input; and

in response to receiving the user input, displaying, on the display, a multi-device interface that includes:

one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and

a media playback status of the plurality of available playback devices.

25. An electronic device, comprising:

a display;

means for receiving a user input; and

means for, responsive to receiving the user input, displaying, on the display, a multi-device interface that includes:

one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and

a media playback status of the plurality of available playback devices.

26. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for performing the method of any of claims 1-22.

27. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 1-22.

28. An electronic device, comprising:

a display; and

means for performing the method of any of claims 1-22.

29. A computer-implemented method, comprising:

at a device with a display:

displaying, on the display, a first current media interface for the device, wherein the first current media interface includes:

a representation of a first media item available for playback on the device, and

a playback device selection affordance;

receiving a first user input representing selection of the playback device selection affordance;

in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device;

receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and

in response to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

30. The method of claim 29, wherein configuring the device to initiate playback on the plurality of available playback devices comprises transmitting an instruction to the available playback devices to initiate playback.

31. The method of claim 29, further comprising, while the device is configured to initiate playback on the plurality of available playback devices:

after ceasing to display the one or more affordances associated with the plurality of available playback devices, receiving a third user input, at the device, representing a request to initiate playback of a second media item;

in response to receiving the third user input, transmitting a request to initiate playback of the second media device on the plurality of available playback devices.

32. The method of any of claims 29-31, wherein the first current media interface for the device includes an indication of a media item currently playing back on the device, the method further comprising:

further in response to receiving the first user input, concurrently displaying, on the display:

a first region that includes the one or more affordances associated with the plurality of available playback devices; and

a second region that includes the indication of the media item currently playing back on the device;

further in response to receiving the second user input, visually updating the second region to display an indication of a media item currently playing back on the plurality of available playback devices.

33. The method of any of claims 29-32, wherein the one or more affordances associated with the plurality of available playback devices connected to the device include information identifying media currently being played back by the plurality of available playback devices.

34. The method of claim 33, wherein the information identifying the media currently being played back by the plurality of available playback devices includes one or more of: a song title, an artist, a movie title, an episode name, and a URL.

35. The method of any of claims 29-34, further comprising, while the device is configured to initiate playback on the plurality of available playback devices:

transmitting an instruction to initiate media playback of a third media item on the plurality of available playback devices,

wherein the plurality of available playback devices are currently playing a fourth media item that was initiated by another device different than the device; and

subsequent to transmitting the instruction to initiate playback of the third media item, receiving an indication that the plurality of devices are currently playing back the third media item.

36. The method of claim 35, further comprising:

determining that the plurality of available playback devices are currently playing content that was initiated by another device;

in accordance with the determination that the plurality of available playback devices are currently playing content that was initiated by another device, requesting user input confirmation that the instruction to initiate media playback of a third media item on the plurality of available playback devices;

receiving a user input response to the request for user input confirmation;

in accordance with the user input response being a positive confirmation, transmitting the instruction to initiate media playback of the third media item on the plurality of available playback devices; and

in accordance with the user input response being a negative confirmation, forgoing transmitting the instruction to initiate media playback of the third media item on the plurality of available playback devices.

37. The method of any of claims 29-36, further comprising:

further in response to receiving the second user input, ceasing displaying the one or more affordances associated with the plurality of available playback devices.

38. The method of any of claims 29-37, wherein a media item includes one or more of audio data, video data, or both.

39. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for:

displaying, on the display, a first current media interface for the device, wherein the first current media interface includes:

a representation of a first media item available for playback on the device, and

a playback device selection affordance;

receiving a first user input representing selection of the playback device selection affordance;

in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device;

receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and

in response to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

40. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

displaying, on the display, a first current media interface for the device, wherein the first current media interface includes:

a representation of a first media item available for playback on the device, and

a playback device selection affordance;

receiving a first user input representing selection of the playback device selection affordance;

in response to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device;

receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and

in response to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

41. An electronic device, comprising:

a display;

means for displaying, on the display, a first current media interface for the device, wherein the first current media interface includes:

a representation of a first media item available for playback on the device, and

a playback device selection affordance;

means for receiving a first user input representing selection of the playback device selection affordance;

means for, responsive to receiving the first user input, displaying, on the display, one or more affordances associated with a plurality of available playback devices connected to the device;

means for receiving a second user input representing selection of the one or more affordances associated with the plurality of available playback devices; and

means for, responsive to receiving the second user input, configuring the device to initiate playback on the plurality of available playback devices.

42. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for performing the method of any of claims 29-38.

43. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 29-38.

44. An electronic device, comprising:

a display; and

means for performing the method of any of claims 29-38.

45. A computer-implemented method, comprising:

at a first device:

detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition,

wherein the first device is connected to the second device via a communication link;

in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link;

subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received;

in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and

in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

46. The method of claim 45, wherein the response audio tone signal represents an indication that the audio tone signal was detected by the second device.

47. The method of any of claims 45-46, wherein at least one of the audio tone signal and the response audio tone signal comprises encoded data.

48. The method of claim 47, wherein the response audio tone signal comprises encoded data, the method further comprising:

further in accordance with a determination that the response audio tone signal from the second device has been received:

converting the encoded data, of the response audio tone signal, into decoded data; and

using at least a portion of the decoded data to log in to a user account of a cloud-based service.

49. The method of any of claims 45-48, wherein the response audio tone signal comprises data that confirms the identity of the second device.

50. The method of any of claims 45-49, wherein the first device includes a memory, the method further comprising:

receiving one or more configuration settings from the second device; and

in response to receiving the one or more configuration settings, storing the configuration settings in the memory of the first device.

51. The method of claim 50, wherein the one or more configuration settings includes a location identifier representing a physical location of the first device.

52. The method of claim 50, wherein the one or more configuration settings includes an indication that the first device is included in a device group, wherein the device group includes the first device and a third device.

53. The method of claim 52, wherein the first and the second device are connected to a local area network via an access point, the method further comprising:

establishing a communication link to the third device, wherein the communication link to the third device does not include the access point; and

synchronizing media playback with the third device using the communication link to the third device.

54. The method of any of claims 45-53, wherein audibly outputting the audio tone signal comprises:

audibly outputting the audio tone signal;

determining whether the second device detected the audio tone signal;

in accordance with a determination that the second device did not detect the audio tone signal, repeating audibly outputting the audio tone signal; and

in accordance with a determination that the second device detected the audio tone signal, forgoing repeating audibly outputting the audio tone signal.

55. The method of claim 54, further comprising:

further in accordance with a determination that the second device did not detect the audio tone signal, determining whether the audio tone signal has be audibly outputted a threshold number of times;

in accordance with a determination that the audio tone signal has been audibly outputted a threshold number of times, audibly outputting a dictated passcode;

in accordance with a determination that the audio tone signal has not been audibly outputted a threshold number of times, forgoing audibly outputting a dictated passcode.

56. The method of any of claims 45-55, wherein the audio tone signal is a first audio tone signal, the method further comprising:

in response to progressing from a first step in the configuration process to a second step in the configuration process, outputting a second audio tone signal,

wherein the second audio tone signal is outputted concurrently with the output of a third audio tone signal outputted by the second device,

wherein the second audio tone signal and the third audio tone signal are harmonics.

57. The method of claim 56, further comprising:

in response to progressing from the second step in the configuration process to a third step in the configuration process, outputting a fourth audio tone signal,

wherein the fourth audio tone signal is outputted concurrently with the output of a fifth audio tone signal outputted by the second device,

wherein the fourth audio tone signal and the fifth audio tone signal are harmonics, and

wherein the fourth audio tone signal is a higher frequency tone than the second audio tone signal.

58. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for:

detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition,

wherein the first device is connected to the second device via a communication link;

in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link;

subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received;

in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and

in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

59. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition,

wherein the first device is connected to the second device via a communication link;

in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link;

subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received;

in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and

in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

60. An electronic device, comprising:

means for detecting an indication that a physical proximity between the first device and a second device satisfies a proximity condition,

wherein the first device is connected to the second device via a communication link;

means for, in accordance with detecting the indication, audibly outputting an audio tone signal, wherein the audio tone signal is out-of-band to the communication link;

means for, subsequent to outputting the audio tone signal, determining whether a response audio tone signal from the second device has been received;

means for, in accordance with a determination that the response audio tone signal from the second device has been received, proceeding to a next step in a configuration process; and

means for, in accordance with a determination that the response audio tone signal from the second device has not been received, forgoing proceeding to the next step in the configuration process.

61. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for performing the method of any of claims 45-57.

62. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 45-57.

63. An electronic device, comprising:

a display; and

means for performing the method of any of claims 45-57.

64. A computer-implemented method, comprising:

at a first device with a display:

detecting, while connected to a second device, an indication that a physical proximity between the first device and the second device satisfies a proximity condition; and

in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

65. The method of claim 64, wherein the indication is detected while the first device is not playing media content and while the second device is not playing media content, and

wherein displaying the media information interface comprises displaying, on the display, a first affordance representing a first media item that was recently played on the second device, and

wherein the method further comprises:

receiving a first user input representing selection of the first affordance; and

in response to receiving the first user input, transmitting an instruction to the second device to initiate playback of the first media item.

66. The method of claim 65, further comprising:

receiving a second user input representing a directional swipe gesture on the first affordance;

in response to receiving the second user input, and while continuing to display the first affordance, displaying, on the display, a second affordance representing a second media item that was recently played on the second device, wherein the second affordance was not displayed prior to receiving the second user input.

67. The method of claim 64, wherein the indication is detected while the first device is playing a third media item and while the second device is playing a fourth media item different than the third media item, and

wherein displaying the media information interface comprises displaying, on the display, a third affordance associated with the third media item currently playing on the first device, and

wherein the method further comprises:

receiving a third user input representing selection of the third affordance; and

in response to receiving the third user input, transmitting an instruction to the second device to initiate playback of the third media item.

68. The method of claim 67, further comprising:

further in response to receiving the third user input, ceasing playing the third media item on the first device.

69. The method of claim 67-68, further comprising:

further in response to receiving the third user input, displaying, on the display, playback control affordances that control the playback of the third media item on the second device.

70. The method of claim 64, wherein the indication is detected while the first device is playing a fifth media item and while the second device is not playing media content, and

wherein displaying the media information interface comprises displaying, on the display, a fourth affordance representing the fifth media item, and

wherein the method further comprises:

receiving a fourth user input;

determining whether the fourth user input represents a directional swipe gesture on the fourth affordance; and

in response to receiving the fourth user input, and in accordance with a determination that the fourth user input represents a directional swipe gesture on the fourth affordance:

transmitting an instruction to the second device to initiate playback of the fifth media item; and

ceasing playback, on the first device, of the fifth media item; and

in response to receiving the fourth user input, and in accordance with a determination that the fourth user input does not represent a directional swipe gesture on the fourth affordance:

forgoing transmitting an instruction to the second device to initiate playback of the fifth media item; and

continue playback, on the first device, of the fifth media item.

71. The method of claim 70, wherein the first device ceases playback of the fifth media item at a playback time of the fifth media item, and

wherein transmitting the instruction to the second device to initiate playback of the fifth media item comprises transmitting an instruction to the second device to begin playback at approximately the playback time of the fifth media item.

72. The method of claim 70, wherein displaying the media information interface comprises displaying, on the display, a fifth affordance;

wherein the method further comprises:

determining whether the fourth user input represents a selection of the fifth affordance; and

in response to receiving the fourth user input:

in accordance with a determination that the fourth user input represents a selection of the fifth affordance, displaying, on the display, a multi-device interface; and

in accordance with a determination that the fourth user input does not represent a selection of the fifth affordance, forgoing displaying the multi-device interface.

73. The method of claim 64, wherein the indication is detected while the first device is not playing media content and while the second device is playing a sixth media item, and

wherein displaying the media information interface comprises displaying, on the display, a sixth affordance representing the sixth media item, and

wherein the method further comprises:

receiving a fifth user input;

determining whether the fifth user input represents a directional swipe gesture on the fifth affordance; and

in response to receiving the fifth user input, and in accordance with a determination that the fifth user input represents a directional swipe gesture on the sixth affordance:

transmitting an instruction to the second device to cease playback of the sixth media item; and

initiating playback, on the first device, of the sixth media item; and

in response to receiving the fifth user input, and in accordance with a determination that the fifth user input does not represent a directional swipe gesture on the sixth affordance:

forgoing transmitting an instruction to the second device to cease playback of the sixth media item; and

forgoing initiating playback, on the first device, of the sixth media item.

74. The method of claim 73, further comprising:

determining whether the fifth user input represents a selection of the sixth affordance; and

further in response to receiving the fifth user input:

in accordance with a determination that the fifth user input represents a selection of the sixth affordance, displaying, on the display, playback control affordances for controlling the playback of the sixth media item on the second device; and

in accordance with a determination that the fifth user input does not represent a selection of the sixth affordance, forgoing displaying, on the display, playback control affordances for controlling the playback of the sixth media item on the second device.

75. The method of claim 73, wherein the second device ceases playback of the sixth media item at a playback time of the sixth media item, and

wherein initiating playback, on the first device, of the sixth media item comprises beginning playback at approximately the playback time of the sixth media item.

76. The method of any of claims 73-75, wherein initiating playback, on the first device, of the sixth media item comprises displaying, on the display:

an identification of the sixth media item; and

playback control affordances for controlling playback of the sixth media item on the first device.

77. The method of any of claims 64-76, wherein the first device is connected to a plurality of devices, wherein the plurality of devices includes the second device, and wherein the displayed media information interface corresponds to the second device.

78. The method of any of claims 64-77, wherein the first device is connected to the second device via a wireless communication link.

79. The method of any of claims 64-78, wherein the media information interface comprises an indication that the media content playing back on the first device or the second device can be transferred to the other device not playing back the media content.

80. The method of any of claims 64-79, wherein a media item is selected from the group consisting of an audio file, an audio stream, a video file, a video stream, a phone call, and a video call.

81. The method of any of claims 64-80, wherein the first device is associated with a first user account, wherein the second device is associated with a set of user accounts, the method further comprising:

further in accordance with detecting that the physical proximity satisfies the proximity condition:

in accordance with a determination that the first user account is a member of the set of user accounts, displaying an indication that media playback can be transferred between the first device and the second device;

in accordance with a determination that the first user account is not a member of the set of user accounts, forgoing displaying the indication that media playback can be transferred between the first device and the second device.

82. The method of any of claims 64-81, further comprising:

in accordance with a determination that the physical proximity between the first device and the second device has satisfied the proximity condition continuously for a threshold amount of time, automatically transferring playback of a media item between the first device and the second device; and

in accordance with a determination that the physical proximity between the first device and the second device has not satisfied the proximity condition continuously for a threshold amount of time, forgoing automatically transferring playback of a media item between the first device and the second device.

83. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, wherein the electronic device is a first device, the one or more programs including instructions for:

detecting, while connected to a second device, an indication that a physical proximity between the first device and the second device satisfies a proximity condition; and

in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

84. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

detecting, while connected to a second device, an indication that a physical proximity between the electronic device and the second device satisfies a proximity condition; and

in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

85. An electronic device, comprising:

a display;

means for detecting, while connected to a second device, an indication that a physical proximity between the electronic device and the second device satisfies a proximity condition; and

means for in accordance with detecting that the physical proximity satisfies the proximity condition, displaying, on the display, a media information interface.

86. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for performing the method of any of claims 64-82.

87. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 64-82.

88. An electronic device, comprising:

a display; and

means for performing the method of any of claims 64-82.

89. A computer-implemented method, comprising:

at a device with a touch-sensitive surface:

detecting a touch on the touch-sensitive surface;

in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media;

in accordance with a determination that the device is currently playing media:

determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping;

in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and

in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and

in accordance with a determination that the device is not currently playing media, initiating playback of media.

90. The method of claim 89, wherein the second area is comprised of a plurality of non-contiguous and non-overlapping areas of the touch-sensitive surface.

91. The method of any of claims 89-90, wherein the second area is comprised of a first sub-area and a second sub-area, and wherein adjusting the volume of the currently playing media comprises:

lowering the volume of the currently playing media if the location of the touch is within the first sub-area; and

raising the volume of the currently playing media if the location of the touch is within the second sub-area.

92. The method of any of claims 89-91, wherein the touch is a first touch, the method further comprising, subsequent to adjusting the volume of the currently playing media in accordance a determination that the location of the touch is within the second area:

temporarily expanding the second area to create an expanded second area for a predetermined amount of time after detecting the first touch on the touch-sensitive surface;

receiving a second touch on the touch-sensitive surface,

wherein the second touch is located within the expanded second area, and

wherein the second touch is not located within the second area;

determining whether the second touch was received during the predetermined amount of time after detecting the first touch;

in accordance with a determination that the second touch was received during the predetermined amount of time after detecting the first touch, adjusting the volume of the currently playing media; and

in accordance with a determination that the second touch was not received during the predetermined amount of time after detecting the first touch, cease playing media.

93. The method of any of claims 89-92, wherein the touch-sensitive surface includes a display, the method further comprising:

further in accordance with a determination that the device is currently playing media, displaying, on the display, an indicator within the second area;

further in accordance with a determination that the device is not currently playing media, forgoing displaying the indicator within the second area.

94. The method of any of claims 89-93, wherein adjusting the volume of the currently playing media comprises:

adjusting a media playback volume; and

adjusting a virtual assistant volume, wherein the virtual assistant volume is proportional to the media playback volume.

95. The method of claim 94, wherein the virtual assistant volume is non-zero when the media playback volume is zero.

96. The method of claim 94, wherein a maximum value of the virtual assistant volume is lower than a maximum value of the media playback volume.

97. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device witha touch-sensitive surface, the one or more programs including instructions for:

detecting a touch on the touch-sensitive surface;

in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media;

in accordance with a determination that the device is currently playing media:

determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping;

in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and

in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and

in accordance with a determination that the device is not currently playing media, initiating playback of media.

98. An electronic device, comprising:

a touch-sensitive surface;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

detecting a touch on the touch-sensitive surface;

in response to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media;

in accordance with a determination that the device is currently playing media:

determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping;

in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and

in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and

in accordance with a determination that the device is not currently playing media, initiating playback of media.

99. An electronic device, comprising:

a touch-sensitive surface;

means for detecting a touch on the touch-sensitive surface;

means for, responsive to detecting the touch on the touch-sensitive surface, determining whether the device is currently playing media;

means for, in accordance with a determination that the device is currently playing media:

determining whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, wherein the first area and the second area are non-overlapping;

in accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, ceasing playing media; and

in accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjusting a volume of the currently playing media; and

means for, in accordance with a determination that the device is not currently playing media, initiating playback of media.

100. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display and a touch-sensitive surface, the one or more programs including instructions for performing the method of any of claims 89-96.

101. An electronic device, comprising:

a display;

a touch-sensitive surface;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 89-96.

102. An electronic device, comprising:

a display;

a touch-sensitive surface; and

means for performing the method of any of claims 89-96.

103. A computer-implemented method, comprising:

at a first device:

receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device;

adding the first media item to a playback queue;

receiving a second request to queue a second media item for playback on the set of devices;

adding the second media item to the playback queue;

accessing the first media item from a first media streaming service;

initiating playback of the first media item on the set of devices;

accessing the second media item from a second media streaming service different than the first media streaming service; and

initiating playback of the second media item on the set of devices.

104. The method of claim 103,

wherein the first request is received from a first requester,

wherein the first request includes identification of the first requester,

wherein the second request is received from a second requester different than the first requester, and

wherein the second request includes identification of the second requester.

105. The method of claim 104,

wherein the first requester is a second device associated with a first user account, and

wherein the second requester is a third device associated with a second user account.

106. The method of claim 104-105, further comprising:

further in response to receiving the first request, determining whether the first requester has permission to add a media item to the playback queue;

in accordance with a determination that the first requester has permission to add a media item to the playback queue, adding the first media item to the playback queue; and

in accordance with a determination that first requester does not have permission to add a media item to the playback queue, forgoing adding the first media item to the playback queue.

107. The method of claim 106, wherein determining whether the first requester has permission to add a media item to the playback queue comprises determining whether a user account associated with the first requester has permission to add a media item to the playback queue, the method further comprising:

in accordance with a determination that the user account associated with the first requester has permission to add a media item to the playback queue, adding the first media item to the playback queue; and

in accordance with a determination that the user account associated with the first requester does not have permission to add a media item to the playback queue, forgoing adding the first media item to the playback queue.

108. The method of claim 104,

receiving, from the second requester, a third request to remove the first media item from the playback queue;

determining whether the second requester has permission to remove the first media item from the playback queue;

in accordance with a determination that the second requester has permission to remove the first media item from the playback queue, removing the first media item from the playback queue; and

in accordance with a determination that the second requester does not have permission to remove the first media item from the playback queue, forgoing removing the first media item from the playback queue.

109. The method of claim 108, wherein determining whether the second requester has permission to remove the first media item from the playback queue comprises:

determining whether the second requester is a member of a predefined set of requesters associated with the first device;

in accordance with a determination that the second requester is a member of a predefined set of requesters associated with the first device, removing the first media item from the playback queue; and

in accordance with a determination that the second requester is a member of a predefined set of requesters associated with the first device, forgoing removing the first media item from the playback queue.

110. The method of any of claims 104-109, further comprising:

transmitting, to a fourth device, an indication of the content of the playback queue for display on the fourth device, wherein the indication of the content of the playback queue includes:

an indication of the first requester that is associated with the first media item in the playback queue, and

an indication of the second requester that is associated with the second media item in the playback queue.

111. The method of any of claims 103-110, further comprising, subsequent to adding the first media item to the playback queue:

determining whether a queue removal condition for the first media item is satisfied;

in accordance with a determination that the queue removal condition for the first media item is satisfied, removing the first media item from the playback queue; and

in accordance with a determination that the queue removal condition for the first media item is not satisfied, forgoing removing the first media item from the playback queue.

112. The method of any of claims 103-111, wherein the first device is not included in the set of devices.

113. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for:

receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device;

adding the first media item to a playback queue;

receiving a second request to queue a second media item for playback on the set of devices;

adding the second media item to the playback queue;

accessing the first media item from a first media streaming service;

initiating playback of the first media item on the set of devices;

accessing the second media item from a second media streaming service different than the first media streaming service; and

initiating playback of the second media item on the set of devices.

114. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device;

adding the first media item to a playback queue;

receiving a second request to queue a second media item for playback on the set of devices;

adding the second media item to the playback queue;

accessing the first media item from a first media streaming service;

initiating playback of the first media item on the set of devices;

accessing the second media item from a second media streaming service different than the first media streaming service; and

initiating playback of the second media item on the set of devices.

115. An electronic device, comprising:

means for receiving a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device;

means for adding the first media item to a playback queue;

means for receiving a second request to queue a second media item for playback on the set of devices;

means for adding the second media item to the playback queue;

means for accessing the first media item from a first media streaming service;

means for initiating playback of the first media item on the set of devices;

means for accessing the second media item from a second media streaming service different than the first media streaming service; and

means for initiating playback of the second media item on the set of devices.

116. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 103-112.

117. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 103-112.

118. An electronic device, comprising:

means for performing the method of any of claims 103-112.

119. A computer-implemented method, comprising:

at a first device:

receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device;

adding the first media item to a playback queue;

receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first;

adding the second media item to the playback queue;

initiating playback of the first media item on the set of devices; and

initiating playback of the second media item on the set of devices.

120. The method of claim 119,

wherein the first request includes a first identifier of at least one of the first user and the second device, and

wherein the second request includes a second identifier of at least one of the second user and the third device.

121. The method of any of claims 119-120,

wherein the second device is associated with a first user account that is associated with the first user, and

wherein the third device is associated with a second user account that is associated with the second user.

122. The method of any of claims 120-121, further comprising:

transmitting, to a fourth device, an indication of the content of the playback queue for display on the fourth device, wherein the indication of the content of the playback queue includes:

the first identifier associated with the first media item in the playback queue, and

the second identifier associated with the second media item in the playback queue.

123. The method of any of claims 119-122, further comprising:

further in response to receiving the second request:

in accordance with a determination that the third device has permission to add a media item to the playback queue, adding the second media item to the playback queue; and

in accordance with a determination that the third device does not have permission to add a media item to the playback queue, forgoing adding the second media item to the playback queue.

124. The method of any of claims 119-123, further comprising:

further in response to receiving the second request:

in accordance with a determination that the second user associated with the third device has permission to add a media item to the playback queue, adding the second media item to the playback queue; and

in accordance with a determination that the second user associated with the third device does not have permission to add a media item to the playback queue, forgoing adding the second media item to the playback queue.

125. The method of claim 119-124, further comprising:

receiving, from the third device, a third request to remove the first media item from the playback queue;

in response to receiving the third request:

in accordance with a determination that the third device has permission to remove the first media item from the playback queue, removing the first media item from the playback queue; and

in accordance with a determination that the third device does not have permission to remove the first media item from the playback queue, forgoing removing the first media item from the playback queue.

126. The method of claim 125, further comprising:

in response to receiving the third request:

in accordance with a determination that the second user is a member of a predefined set of users associated with the first device, removing the first media item from the playback queue; and

in accordance with a determination that the second user is not a member of a predefined set of users associated with the first device, forgoing removing the first media item from the playback queue.

127. The method of any of claims 119-126, further comprising, subsequent to adding the first media item to the playback queue:

in accordance with a determination that a queue removal condition for the first media item is satisfied, removing the first media item from the playback queue; and

in accordance with a determination that the queue removal condition for the first media item is not satisfied, forgoing removing the first media item from the playback queue.

128. The method of any of claims 119-127, wherein the first device is not included in the set of devices.

129. The method of any of claims 119-128, wherein the first device is associated with a media preference profile, wherein the media preference profile is associated with the first user, the method further comprising:

further in response to receiving, from the second device associated with the first user, the request to queue the first media item for playback, updating the media preference profile based on the first media item; and

further in response to receiving, from the third device associated with the second user, the request to queue the second media item for playback, forgoing updating the media preference profile based on the second media item.

130. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for:

receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the first device;

adding the first media item to a playback queue;

receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first;

adding the second media item to the playback queue;

initiating playback of the first media item on the set of devices; and

initiating playback of the second media item on the set of devices.

131. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device;

adding the first media item to a playback queue;

receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first;

adding the second media item to the playback queue;

initiating playback of the first media item on the set of devices; and

initiating playback of the second media item on the set of devices.

132. An electronic device, comprising:

means for receiving, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, wherein the set of devices includes the electronic device;

means for adding the first media item to a playback queue;

means for receiving, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, wherein the second user is different than the first;

means for adding the second media item to the playback queue;

means for initiating playback of the first media item on the set of devices; and

means for initiating playback of the second media item on the set of devices.

133. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 119-129.

134. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 119-129.

135. An electronic device, comprising:

means for performing the method of any of claims 119-129.

136. A computer-implemented method, comprising:

at a device:

while outputting an audible signal from the device, detecting a user input voice command from a voice command source;

in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and

while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

137. The method of claim 136, wherein the device is a first device, and wherein the audible signal is a first audible signal, the method further comprising:

determining that a second device is currently outputting a second audible signal; and

wherein adjusting, by the first device, the output of the first audible signal is further based on a known influence that audio output by the first device has on the second device.

138. The method of any of claims 136-137, wherein the device includes a speaker array, wherein each speaker in the speaker array receives a respective output signal, and wherein adjusting the output of the audible signal based comprises:

performing an audio processing technique using at least:

the respective output signals of each speaker in the speaker array, and

the spatial position of the voice command source; and

adjusting at least one of the respective output signals corresponding to a speaker of the speaker array.

139. The method of any of claims 136-138, wherein the spatial position is a first spatial position, and

wherein adjusting the output of the audible signal further comprises outputting the audible signal such that it will be perceived, by a listener at the first spatial position, to emanate from a second spatial position that is in a direction, relative to the device, away from the first spatial position of the voice command source.

140. The method of any of claims 136-139, further comprising, subsequent to adjusting the output of the audible signal:

detecting that the user input voice command from the voice command source has ended; and

in response to detecting that the user input voice command from the voice command source has ended, ceasing adjusting the output the audible signal.

141. The method of any of claims 136-140, wherein determining the spatial position of the voice command source relative to the device comprises:

determining a distance of the voice command source relative to the device; and

determining an angular position of the voice command source relative to the device.

142. The method of any of claims 136-141, wherein the audible signal is a first audible signal, the method further comprising, while adjusting the output of the audible signal:

outputting a second audio signal based on the spatial position of the voice command source.

143. The method of claim 142, wherein the spatial position is a first spatial position, and

wherein outputting the second audible signal comprises outputting the second audible signal such that it will be perceived, by a listener at the first spatial position, to emanate from a third spatial position that is in a direction, relative to the device, toward the first spatial position of the voice command source.

144. The method of any of claims 142-143, wherein the first audible signal represents playback of a media item, and wherein the second audible signal represents a virtual assistant output.

145. The method of any of claims 136-144, wherein the device has a plurality of microphones.

146. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for:

while outputting an audible signal from the device, detecting a user input voice command from a voice command source;

in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and

while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

147. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

while outputting an audible signal from the device, detecting a user input voice command from a voice command source;

in response to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and

while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

148. An electronic device, comprising:

means for, while outputting an audible signal from the device, detecting a user input voice command from a voice command source;

means for, responsive to detecting the user input voice command, determining a spatial position of the voice command source relative to the device; and

means for, while continuing to output the audible signal, adjusting the output of the audible signal based on the spatial position of the voice command source.

149. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 136-145.

150. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 136-145.

151. An electronic device, comprising:

means for performing the method of any of claims 136-145.

152. A computer-implemented method, comprising:

at a first device associated with a first user account, wherein the first user account is associated with a first media preference profile:

detecting that a second device associated with a second user account is connected to the first device,

wherein the second user account is associated with a second media preference profile different than the first media preference profile; and

initiating playback of media based on the first media preference profile and the second media preference profile.

153. The method of claim 152, further comprising:

determining a combined preference profile based on the first media preference profile and the second media preference profile.

154. The method of claim 153, wherein determining a combined preference profile based on the first media preference profile and the second media preference profile comprises:

determining a media preference that is related to both the first media preference profile and the second media preference profile.

155. The method of any of claims 152-154, further comprising, while continuing to detect that the second device is connected to the first device:

adding a plurality of media items to a playback queue based on the first media preference profile and the second media preference profile; and

initiating playback of the plurality of media items.

156. The method of any of claims 152-155, further comprising:

receiving a request to initiate playback on the first device; and

in response to receiving the request, initiating the playback of media based on the first media preference profile and the second media preference profile.

157. The method of claim 156, wherein the request is received from the second device.

158. The method of claim 156, wherein the first device includes a display, and wherein the method further comprises:

displaying, on the display, an affordance; and

wherein receiving the request to initiate playback on the first device comprises receiving user input selection of the affordance; and

in response to receiving user input selection of the affordance, initiating the playback of media based on the first media preference profile and the second media preference profile.

159. The method of any of claims 152-158, wherein initiating playback of media based on the first media preference profile and the second media preference profile comprises:

transmitting an instruction to a third device to initiate playback.

160. The method of any of claims 152-159, further comprising:

receiving a media item;

in response to receiving the media item:

in accordance with a determination that the media item was received from the first device, updating the first media preference profile based on the media item; and

in accordance with a determination that the media item was received from the second device, forgoing updating the first media preference profile based on the media item.

161. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device associated with a first user account, wherein the first user account is associated with a first media preference profile, the one or more programs including instructions for:

detecting that a second device associated with a second user account is connected to the first device,

wherein the second user account is associated with a second media preference profile different than the first media preference profile; and

initiating playback of media based on the first media preference profile and the second media preference profile.

162. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

detecting that a second device associated with a second user account is connected to the electronic device,

wherein the electronic device is associated with a first user account, wherein the first user account is associated with a first media preference profile:

wherein the second user account is associated with a second media preference profile different than the first media preference profile; and

initiating playback of media based on the first media preference profile and the second media preference profile.

163. An electronic device, comprising:

means for detecting that a second device associated with a second user account is connected to the electronic device,

wherein the electronic device is associated with a first user account, wherein the first user account is associated with a first media preference profile:

wherein the second user account is associated with a second media preference profile different than the first media preference profile; and

means for initiating playback of media based on the first media preference profile and the second media preference profile.

164. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 152-160.

165. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 152-160.

166. An electronic device, comprising:

means for performing the method of any of claims 152-160.

167. A computer-implemented method, comprising:

at a first device with a display:

detecting an audio tone signal indicative of a pairing handshake with a second device,

wherein the first device is connected to the second device via a communication link, and

wherein the audio tone signal is out-of-band to the communication link;

in accordance with detecting the audio tone signal, initiating, on the first device, a configuration process for configuring the second device;

during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location;

receiving user input selection of a location; and

in response to receiving the user input, configuring the second device to be associated with the location.

168. The method of claim 167, wherein the audio tone signal is generated by the second device.

169. The method of any of claims 167-168, wherein the audio tone signal includes an audio passcode.

170. The method of any of claims 167-169, wherein the audio tone signal is received while the first device is physically positioned within a threshold physical proximity to the second device.

171. The method of any of claims 167-170, wherein the audio tone signal is a first audio tone signal, the method further comprising, during the configuration process:

detecting a second audio tone signal comprising encoded data;

in response to detecting the second audio tone signal:

transmitting an acknowledgement to the second device; and

proceeding to a next step in the configuration process.

172. The method of any of claims 167-171, further comprising:

further in response to receiving the user input selection of the location:

in accordance with a determination that a third device is currently associated with the location, displaying, on the display, a configuration user interface that includes a prompt to select whether to create a device group that includes the second device and the third device; and

in accordance with a determination that a third device is not currently associated with the location, forgoing displaying, on the display, the configuration user interface that includes the prompt to select whether to create the device group.

173. The method of any of claims 167-172, further comprising:

receiving user input representing an affirmative request create the device group;

in response to the user input representing an affirmative request create the device group, displaying, on the display, a configuration user interface that includes a prompt to select an audio output channel to associate with the second device;

receiving user input representing selection of an audio output channel; and

in response to receiving the user input representing selection of a selected audio output channel, associating the second device with the selected audio channel.

174. The method of any of claims 167-173, wherein the prompt to select a location includes a plurality of predetermined location identifiers.

175. The method of claim 174, wherein the plurality of predetermined location identifiers includes a user-defined location identifier.

176. The method of any of claims 167-173, wherein the prompt to select a location includes a field for creating a new location identifier, the method further comprising:

receiving user input associated with the field for creating a new location identifier, wherein the user input includes a location identifier;

creating a new location identifier based on the received user input associated with the field for creating the new location identifier; and

associating the second device with the new location identifier.

177. The method of any of claims 167-176, further comprising, during the configuration process:

displaying, on the display, a configuration user interface that includes a prompt requesting permission to transmit user account information to the second device;

receiving user input providing permission to transmit the user account information to the second device; and

in response to receiving the user input providing permission, transmitting the user account information to the second device.

178. The method of claim 177, further comprising, prior to transmitting the user account information to the second device:

in accordance with a determination that the first device is not currently logged into a user account associated with the user account information,

displaying, on the display, a prompt to log in to the user account; and

forgoing transmitting the user account information to the second device; and

in accordance with a determination that the first device is currently logged into the user account associated with the user account information, transmitting the user account information to the second device.

179. The method of any of claims 167-177, further comprising:

in response to progressing from a first step in the configuration process to a second step in the configuration process, outputting a third audio tone signal,

wherein the third audio tone signal is outputted concurrently with the output of a fourth audio tone signal outputted by the second device,

wherein the third audio tone signal and the fourth audio tone signal are harmonics.

180. The method of claim 179, further comprising:

in response to progressing from the second step in the configuration process to a third step in the configuration process, outputting a fifth audio tone signal,

wherein the fifth audio tone signal is outputted concurrently with the output of a sixth audio tone signal outputted by the second device,

wherein the fifth audio tone signal and the sixth audio tone signal are harmonics, and

wherein the fifth audio tone signal is a higher frequency tone than the third audio tone signal.

181. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, wherein the electronic device is a first device, the one or more programs including instructions for:

detecting an audio tone signal indicative of a pairing handshake with a second device,

wherein the first device is connected to the second device via a communication link, and

wherein the audio tone signal is out-of-band to the communication link;

in accordance with detecting the audio tone signal, initiating, on the first device, a configuration process for configuring the second device;

during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location;

receiving user input selection of a location; and

in response to receiving the user input, configuring the second device to be associated with the location.

182. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

detecting an audio tone signal indicative of a pairing handshake with a second device,

wherein the electronic device is connected to the second device via a communication link, and

wherein the audio tone signal is out-of-band to the communication link;

in accordance with detecting the audio tone signal, initiating, on the electronic device, a configuration process for configuring the second device;

during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location;

receiving user input selection of a location; and

in response to receiving the user input, configuring the second device to be associated with the location.

183. An electronic device, comprising:

a display;

means for detecting an audio tone signal indicative of a pairing handshake with a second device,

wherein the electronic device is connected to the second device via a communication link, and

wherein the audio tone signal is out-of-band to the communication link;

means for, in accordance with detecting the audio tone signal, initiating, on the electronic device, a configuration process for configuring the second device;

means for, during the configuration process, displaying, on the display, a configuration user interface that includes a prompt to select a location;

means for receiving user input selection of a location; and

means for, responsive to receiving the user input, configuring the second device to be associated with the location.

184. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device with a display, the one or more programs including instructions for performing the method of any of claims 167-180.

185. An electronic device, comprising:

a display;

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 167-180.

186. An electronic device, comprising:

a display; and

means for performing the method of any of claims 167-180.

187. A computer-implemented method, comprising:

at a first device:

receiving a request, from a third device, to initiate media playback on at least one device of a set of devices,

wherein the set of devices comprises the first device and a second device,

wherein the third device is not a member of the set of devices, and

wherein the third device is connected to the first device;

in response to receiving the request:

in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and

in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices:

transmitting a request for permission to the second device; and

in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

188. The method of claim 187, further comprising:

in accordance with a determination that a first user account associated with the third device has permission to initiate media playback on the set of devices, determining that the third device has permission, wherein the first user account is not associated with any device of the set of devices; and

in accordance with a determination that the first user account associated with the third device does not have permission to initiate media playback on the set of devices, forgoing determining that the third device has permission.

189. The method of any of claims 187-188, further comprising:

further in accordance with a determination that the third device does not have permission, transmitting the request for permission to a plurality of devices that includes the second device.

190. The method of any of claims 187-189, wherein initiating media playback on the at least one device of the set of devices comprises:

initiating playback on the first device; and

forgoing initiating playback on the second device.

191. The method of any of claims 187-190, further comprising, subsequent to receiving permission from the second device:

receiving a request, from the third device, to initiate playback on a fourth device,

wherein the set of devices comprises the fourth device, and

wherein the fourth device is not included in the at least one device of the set of devices; and

in response to receiving the request to initiate playback on a fourth device, initiating playback on the fourth device.

192. The method of any of claims 187-191, further comprising, subsequent to receiving permission from the second device:

receiving a request, from the third device, to remove a media item from a playback queue on the at least one device of the set of devices;

in response to receiving the request:

in accordance with a determination that the media item was added to the playback queue by the third device, removing the media item from the playback queue; and

in accordance with a determination that that the media item was not added to the playback queue by the third device, forgoing removing the media item from the playback queue.

193. The method of any of claims 187-192, wherein the first device is associated with a media preference profile, the method further comprising:

further in response to receiving the request from the third device that is not a member of the set of devices, forgoing updating the media preference profile based on the request.

194. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device, the one or more programs including instructions for:

receiving a request, from a third device, to initiate media playback on at least one device of a set of devices,

wherein the set of devices comprises the first device and a second device,

wherein the third device is not a member of the set of devices, and

wherein the third device is connected to the first device;

in response to receiving the request:

in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and

in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices:

transmitting a request for permission to the second device; and

in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

195. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving a request, from a third device, to initiate media playback on at least one device of a set of devices,

wherein the set of devices comprises the electronic device and a second device,

wherein the third device is not a member of the set of devices, and

wherein the third device is connected to the electronic device;

in response to receiving the request:

in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and

in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices:

transmitting a request for permission to the second device; and

in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

196. An electronic device, comprising:

means for receiving a request, from a third device, to initiate media playback on at least one device of a set of devices,

wherein the set of devices comprises the electronic device and a second device,

wherein the third device is not a member of the set of devices, and

wherein the third device is connected to the electronic device;

means for, responsive to receiving the request:

in accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiating media playback on the set of devices; and

in accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices:

transmitting a request for permission to the second device; and

in response to receiving permission from the second device, initiating media playback on the at least one device of the set of devices.

197. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 187-193.

198. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 187-193.

199. An electronic device, comprising:

means for performing the method of any of claims 187-193.

200. A computer-implemented method, comprising:

at a first device associated with a user account:

receiving user input representing a request to access data from the user account;

in response to receiving the user input, determining whether a physical positioning of the first device and a second device associated with the user account satisfies a permissive condition;

in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the first device; and

in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the first device.

201. The method of claim 200, wherein providing the data from the user account through the first device comprises:

outputting an audio signal representative of the data from the user account.

202. The method of claim 201, wherein outputting an audio signal representative of the data from the user account comprises:

causing audible output of a dictation representative of the data from the user account using a text-to-speech algorithm.

203. The method of claim 202, wherein the data from the user account comprises data from one or more of: a calendar entry, a notification, an email, a text message, a stored text document, stored contact information, phone call audio, video call audio or video, a voice message, an address, and a media item.

204. The method of any of claims 200-203, wherein determining whether the physical positioning of the first device and second device satisfies the permissive condition comprises:

determining whether a physical proximity between the first device and the second device satisfies a proximity condition.

205. The method of any of claims 200-204, wherein determining whether the physical positioning of the first device and second device satisfies the permissive condition comprises:

determining whether a physical orientation of the first device satisfies an orientation condition.

206. The method of any of claims 200-205, further comprising:

further in accordance with a determination that the physical positioning does not satisfy the permissive condition, outputting an indication that the permissive condition is not satisfied.

207. The method of claim 206, wherein outputting the indication that the permissive condition is not satisfied comprises:

outputting an audible indication of a user action required to satisfy the permissive condition.

208. The method of any of claims 200-207, wherein the user input representing a request to access data from the user account is a first user input, the method further comprising, subsequent to providing the data from the user account through the first device:

receiving a second user input representing a request to access additional data from the user account;

in response to receiving the second user input, determining whether the physical positioning of the first device and a second device associated with the user account satisfies the permissive condition;

in accordance with a determination that the physical positioning satisfies the permissive condition, providing the additional data from the user account through the first device;

in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the additional data from the user account through the first device.

209. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first electronic device associated with a user account, the one or more programs including instructions for:

receiving user input representing a request to access data from the user account;

in response to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition;

in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and

in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

210. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving user input representing a request to access data from a user account, wherein the electronic device is associated with the user account;

in response to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition;

in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and

in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

211. An electronic device, comprising:

means for receiving user input representing a request to access data from a user account, wherein the electronic device is associated with the user account;

means for, responsive to receiving the user input, determining whether a physical positioning of the electronic device and a second device associated with the user account satisfies a permissive condition;

means for, in accordance with a determination that the physical positioning satisfies the permissive condition, providing the data from the user account through the electronic device; and

means for, in accordance with a determination that the physical positioning does not satisfy the permissive condition, forgoing providing the data from the user account through the electronic device.

212. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 200-208.

213. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 200-208.

214. An electronic device, comprising:

means for performing the method of any of claims 200-208.

215. A computer-implemented method, comprising:

at a first device:

receiving a command to initiate playback of media on a set of devices comprising the first device and a second device;

in response to receiving the command, initiating synchronized playback of the media on the set of devices,

wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and

wherein the first device is configured to be a designated heir to the primary device.

216. The method of claim 215, wherein the set of devices further comprises a third device, the method further comprising:

during playback on the set of devices, detecting that the second device is no longer available for playback of the media; and

in response to detecting that the second device is no longer available to playback the media:

determining that the first device is designated as the primary device configured to coordinate the playback of the media on the set of devices;

continuing playback of the media on the first device; and

coordinating the playback of the media on the third device.

217. The method of any of claims 215-216, wherein coordinating the playback of the media on the third device comprises one or more of:

providing a link to the media;

providing a stream of the media; and

providing a copy of the media.

218. The method of any of claims 215-217, further comprising:

receiving a command to cease playback of media on the first device;

in response to receiving the command to cease playback of media on the first device:

ceasing playback of media on the first device; and

designating the third device to be the primary device configured to coordinate the playback of the media on the set of devices.

219. The method of claim 215, wherein the set of devices further comprises a third device, the method further comprising:

during playback on the set of devices, detecting that the second device is no longer available for playback of the media; and

in response to detecting that the second device is no longer available to playback the media:

determining that the third device is designated as the primary device configured to coordinate the playback of the media on the set of devices; and

continuing playback of the media on the first device; and

receiving, from the third device, an instruction for coordinating playback of the media on the first device.

220. The method of any of claims 215-219, further comprising:

receiving, from a fourth device, the command to initiate playback of media on the set of devices; and

receiving, from the fourth device, a designation that the first device is the designated heir.

221. The method of claim 220, wherein the fourth device is not a member of the set of devices.

222. The method of any of claims 215-221, wherein initiating synchronized playback of the media on the set of devices comprises transmitting a copy of a playback queue to each device in the set of devices.

223. The method of any of claims 215-222, further comprising:

during synchronized playback, receiving, at the first device, a playback control command;

in response to the playback control command:

in accordance with the first device being designated the primary device, transmitting an instruction to each of the devices in the set of devices based on the playback control command; and

in accordance with the first device not being designated the primary device, transmitting the playback control command to a designated primary device.

224. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic, wherein the electronic device is a first device, the one or more programs including instructions for:

receiving a command to initiate playback of media on a set of devices comprising the first device and a second device;

in response to receiving the command, initiating synchronized playback of the media on the set of devices,

wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and

wherein the first device is configured to be a designated heir to the primary device.

225. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving a command to initiate playback of media on a set of devices comprising the electronic device and a second device;

in response to receiving the command, initiating synchronized playback of the media on the set of devices,

wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and

wherein the electronic device is configured to be a designated heir to the primary device.

226. An electronic device, comprising:

means for receiving a command to initiate playback of media on a set of devices comprising the electronic device and a second device;

means for, responsive to receiving the command, initiating synchronized playback of the media on the set of devices,

wherein the second device is designated as a primary device configured to coordinate the playback of the media on the set of devices, and

wherein the electronic device is configured to be a designated heir to the primary device.

227. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 215-223.

228. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 215-223.

229. An electronic device, comprising:

means for performing the method of any of claims 215-223.

230. A computer-implemented method, comprising:

at a first device of a set of devices associated with a first user account:

receiving an audible output trigger event;

determining that a first user, associated with the first user account, is located within an audible output range of a device of the set of devices;

in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and

in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

231. The method of claim 230, wherein the device of the set of devices is a second device, and wherein outputting the audible output comprises transmitting an instruction to the second device to output the audible output.

232. The method of any of claims 230-231, wherein outputting the audible output includes playing an audible tone.

233. The method of any of claims 230-232, wherein receiving the audible output trigger event comprises receiving user input representing configuration of the audible output trigger event.

234. The method of any of claims 230-233, wherein receiving the audible output trigger event comprises receiving data representing audible output settings associated with the first user account.

235. The method of any of claims 230-234, wherein the audible output range of the first device is based on one or more of:

a detected physical proximity of the first device to the first user;

a detected presence of the first user in the same room as the first device; and

the output characteristics of the first device.

236. The method of any of claims 230-235, wherein determining that the first user is located within the audible output range of the device of the set of devices comprises one or more of the following:

detecting, at the device, a voice associated with the first user;

detecting, at the device, a third device associated with the user, wherein the third device is not a member of the set of devices; and

receiving a signal, from an external sensor, representative of the first user being physically near the device.

237. The method of any of claims 230-236, further comprising, subsequent to outputting the audible output:

in accordance with a determination that the first user is within an audible output range of a fourth device that is not a member of the set of devices, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

238. The method of claim 237, further comprising:

playing back media content; and

further in accordance with a determination that the first user is within the audible output range of the fourth device that is not a member of the set of devices, ceasing playback of the media content.

239. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, wherein the electronic device is a first device of a set of devices associated with a first user account, the one or more programs including instructions for:

receiving an audible output trigger event;

determining that a first user, associated with the first user account, is located within an audible output range of a device of the set of devices;

in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and

in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

240. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving an audible output trigger event;

determining that a first user, associated with a first user account, is located within an audible output range of a device of a set of devices associated with the first user account, wherein the electronic device is a first device of the set of devices;

in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and

in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

241. An electronic device, comprising:

means for receiving an audible output trigger event;

means for determining that a first user, associated with a first user account, is located within an audible output range of a device of a set of devices associated with the first user account, wherein the electronic device is a first device of the set of devices;

means for, in accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, outputting an audible output; and

means for, in accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgoing outputting the audible output.

242. A non-transitory computer-readable storage medium storing one or more programs configured to be executed by one or more processors of an electronic device, the one or more programs including instructions for performing the method of any of claims 230-238.

243. An electronic device, comprising:

one or more processors; and

memory storing one or more programs configured to be executed by the one or more processors, the one or more programs including instructions for performing the method of any of claims 230-238.

244. An electronic device, comprising:

means for performing the method of any of claims 230-238.

ABSTRACT

The present disclosure generally relates to interfaces and techniques for media playback on one or more devices. In accordance with some embodiments, an electronic device includes a display, one or more processors, and memory. The electronic device receives user input and, in response to receiving the user input, displays, on the display, a multi-device interface that includes: one or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device, and a media playback status of the plurality of available playback devices.

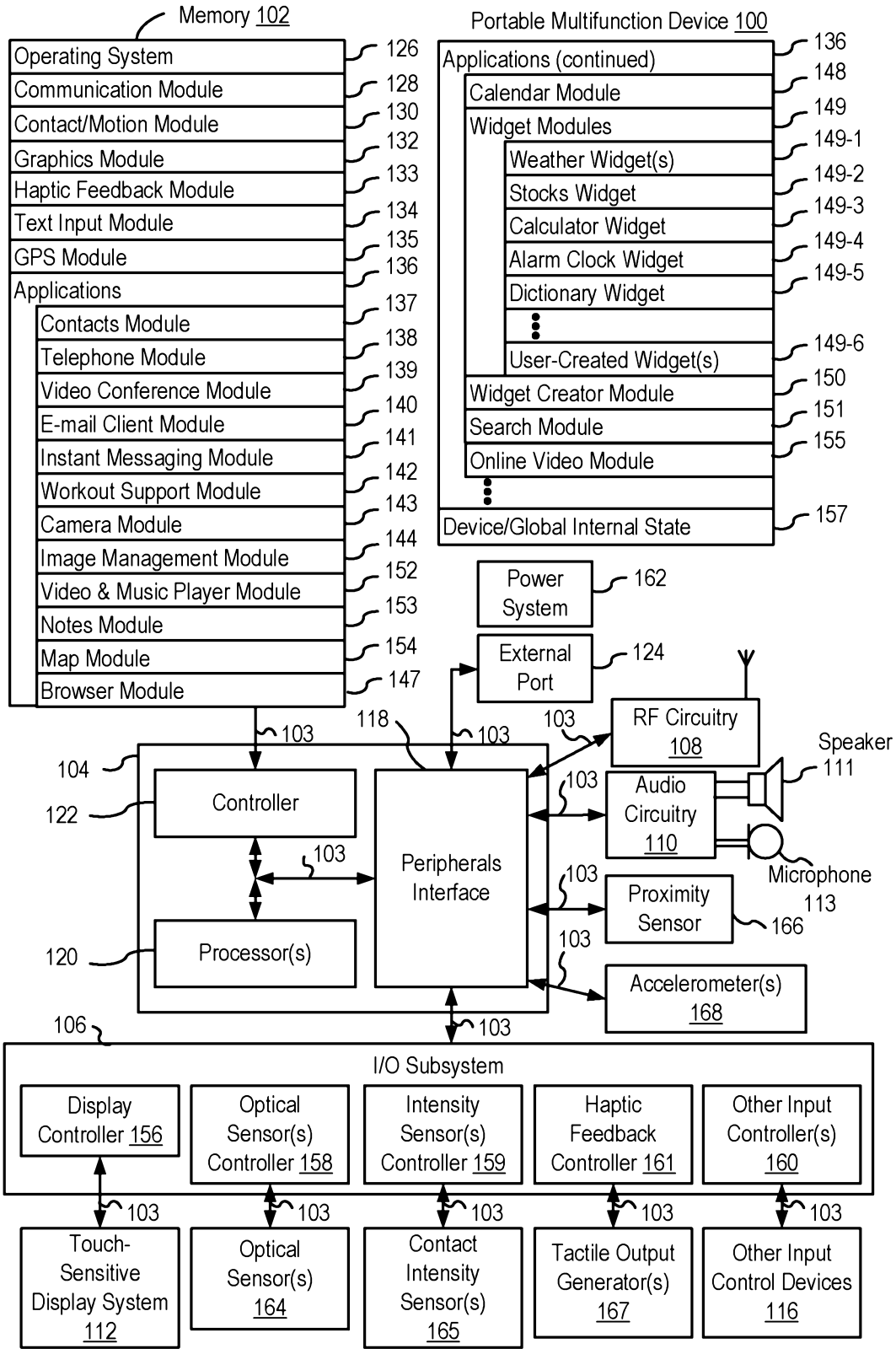

Memory 102
Operating System
126
Communication Module
128
Contact/Motion Module
130
Graphics Module
132
Haptic Feedback Module
133
Text Input Module
134
GPS Module
135
Applications
136
Contacts Module
137
Telephone Module
138
Video Conference Module
139
E-mail Client Module
140
Instant Messaging Module
141
Workout Support Module
142
Camera Module
143
Image Management Module
144
Video & Music Player Module
152
Notes Module
153
Map Module
154
Browser Module
147
Portable Multifunction Device 100
Applications (continued)
136
Calendar Module
148
Widget Modules
149
Weather Widget(s)
149-1
Stocks Widget
149-2
Calculator Widget
149-3
Alarm Clock Widget
149-4
Dictionary Widget
149-5
User-Created Widget(s)
149-6
Widget Creator Module
150
Search Module
151
Online Video Module
155
Device/Global Internal State
157
Power System
162
External Port
124
RF Circuitry 108
Speaker 111
Audio Circuitry 110
Microphone 113
Proximity Sensor
166
Accelerometer(s) 168
Peripherals Interface
Controller
Processor(s)
104
122
120
118
103
I/O Subsystem
106
Display Controller 156
Optical Sensor(s) Controller 158
Intensity Sensor(s) Controller 159
Haptic Feedback Controller 161
Other Input Controller(s) 160
Touch-Sensitive Display System 112
Optical Sensor(s) 164
Contact Intensity Sensor(s) 165
Tactile Output Generator(s) 167
Other Input Control Devices 116

FIG. 1A

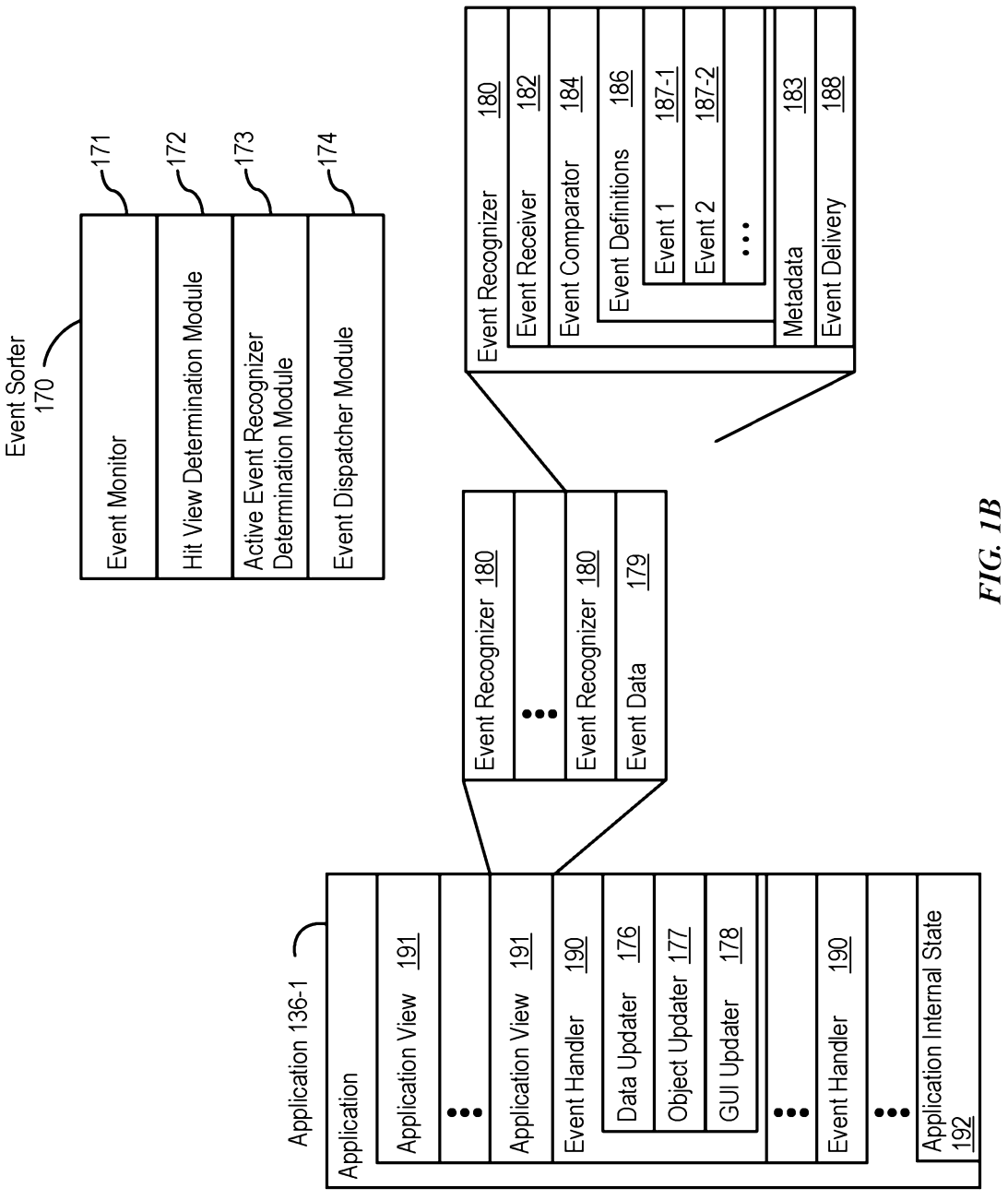

Event Sorter
170
171
Event Monitor
172
Hit View Determination Module
173
Active Event Recognizer
Determination Module
174
Event Dispatcher Module
Event Recognizer 180
Event Receiver 182
Event Comparator 184
Event Definitions 186
Event 1 187-1
Event 2 187-2
...
Metadata 183
Event Delivery 188
Event Recognizer 180
...
Event Recognizer 180
Event Data 179
Application 136-1
Application
Application View 191
...
Application View 191
Event Handler 190
Data Updater 176
Object Updater 177
GUI Updater 178
...
Event Handler 190
...
Application Internal State
192

FIG. 1B

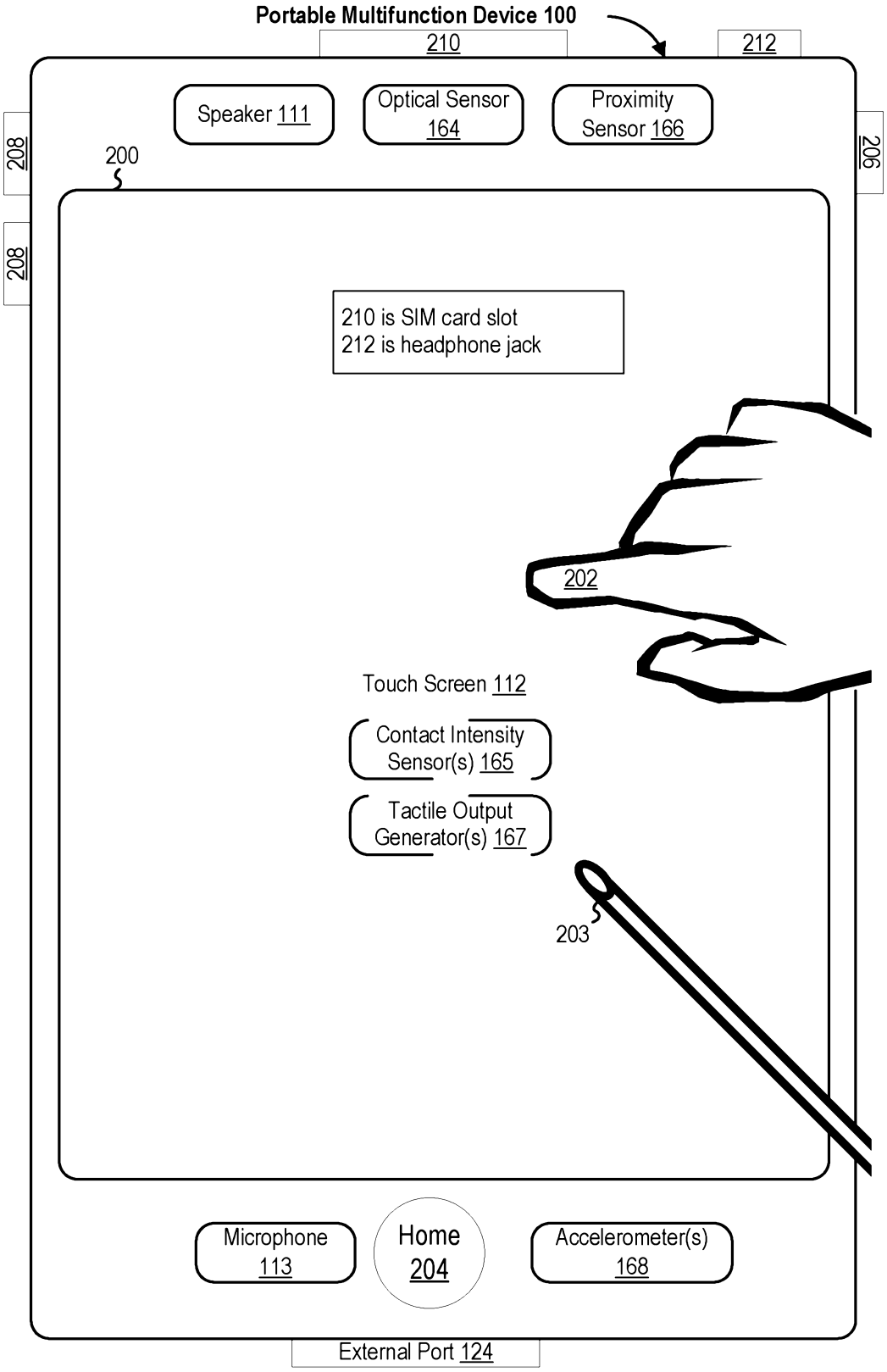

Portable Multifunction Device 100
210
212
Speaker 111
Optical Sensor 164
Proximity Sensor 166
208
208
206
200
210 is SIM card slot
212 is headphone jack
202
Touch Screen 112
Contact Intensity Sensor(s) 165
Tactile Output Generator(s) 167
203
Microphone 113
Home 204
Accelerometer(s) 168
External Port 124

FIG. 2

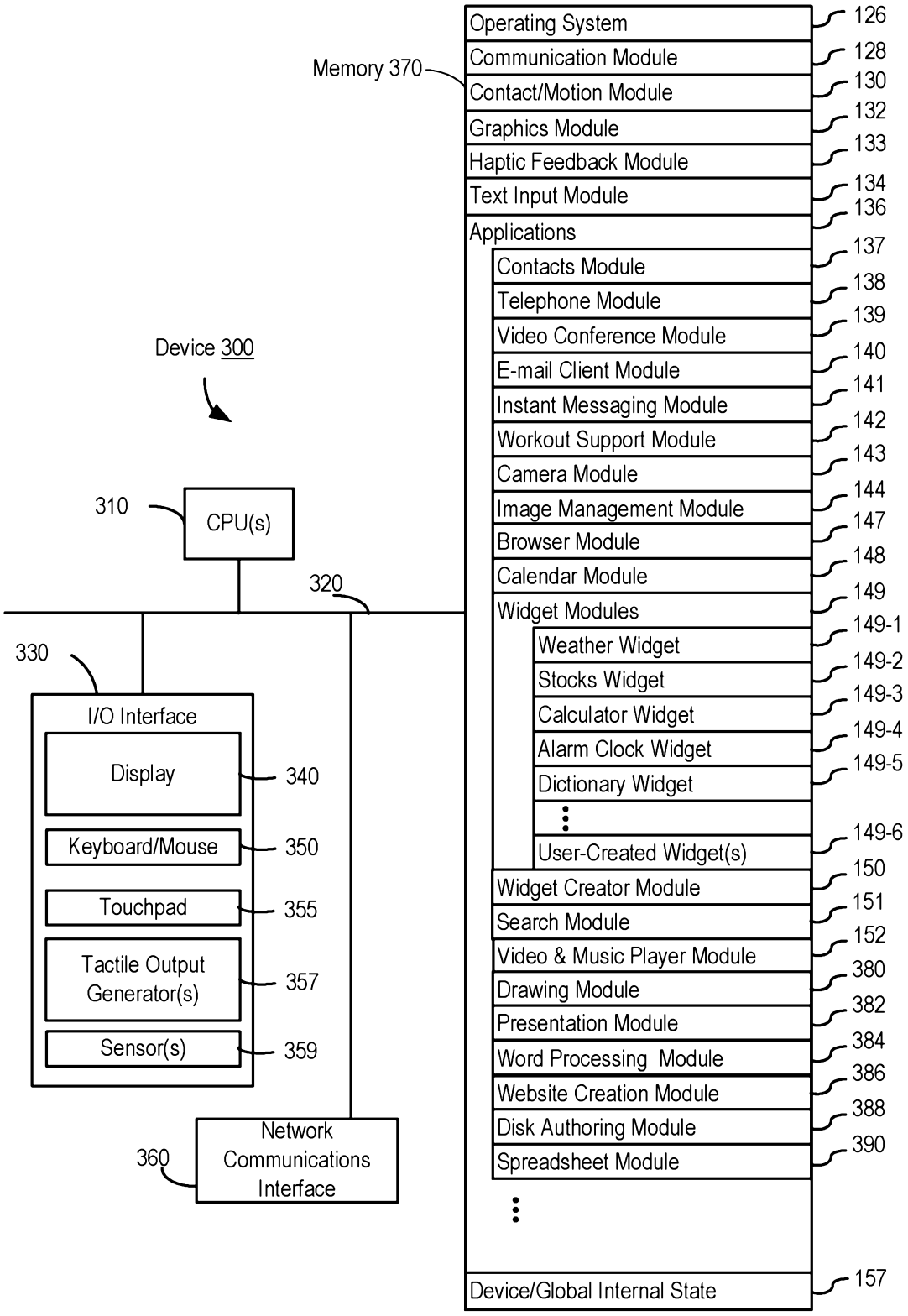

Memory 370
Operating System
126
Communication Module
128
Contact/Motion Module
130
Graphics Module
132
Haptic Feedback Module
133
Text Input Module
134
Applications
136
Contacts Module
137
Telephone Module
138
Video Conference Module
139
E-mail Client Module
140
Instant Messaging Module
141
Workout Support Module
142
Camera Module
143
Image Management Module
144
Browser Module
147
Calendar Module
148
Widget Modules
149
Weather Widget
149-1
Stocks Widget
149-2
Calculator Widget
149-3
Alarm Clock Widget
149-4
Dictionary Widget
149-5
User-Created Widget(s)
149-6
Widget Creator Module
150
Search Module
151
Video & Music Player Module
152
Drawing Module
380
Presentation Module
382
Word Processing Module
384
Website Creation Module
386
Disk Authoring Module
388
Spreadsheet Module
390
Device/Global Internal State
157
Device 300
310
CPU(s)
320
330
I/O Interface
Display
340
Keyboard/Mouse
350
Touchpad
355
Tactile Output Generator(s)
357
Sensor(s)
359
360
Network Communications Interface

FIG. 3

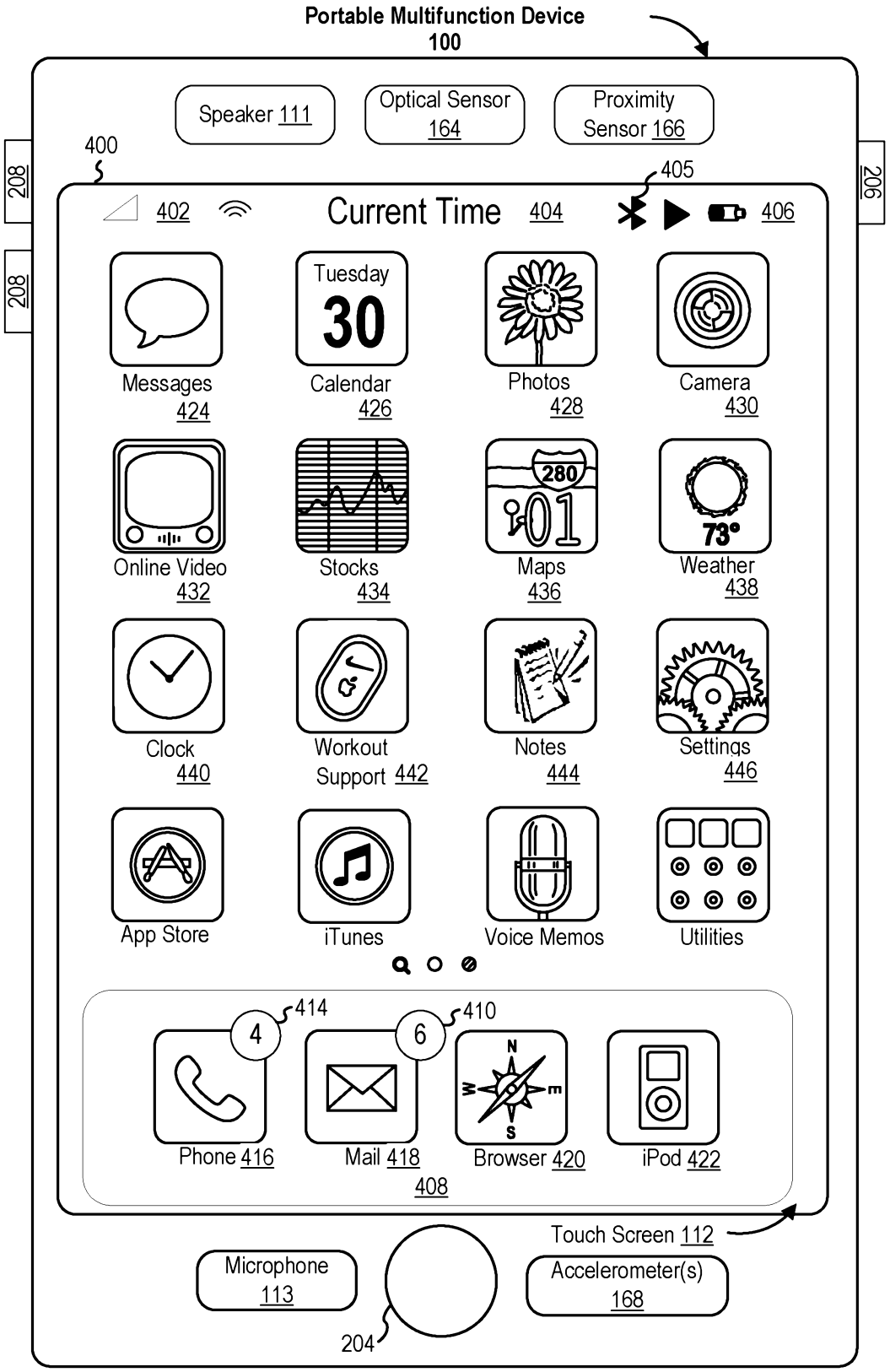
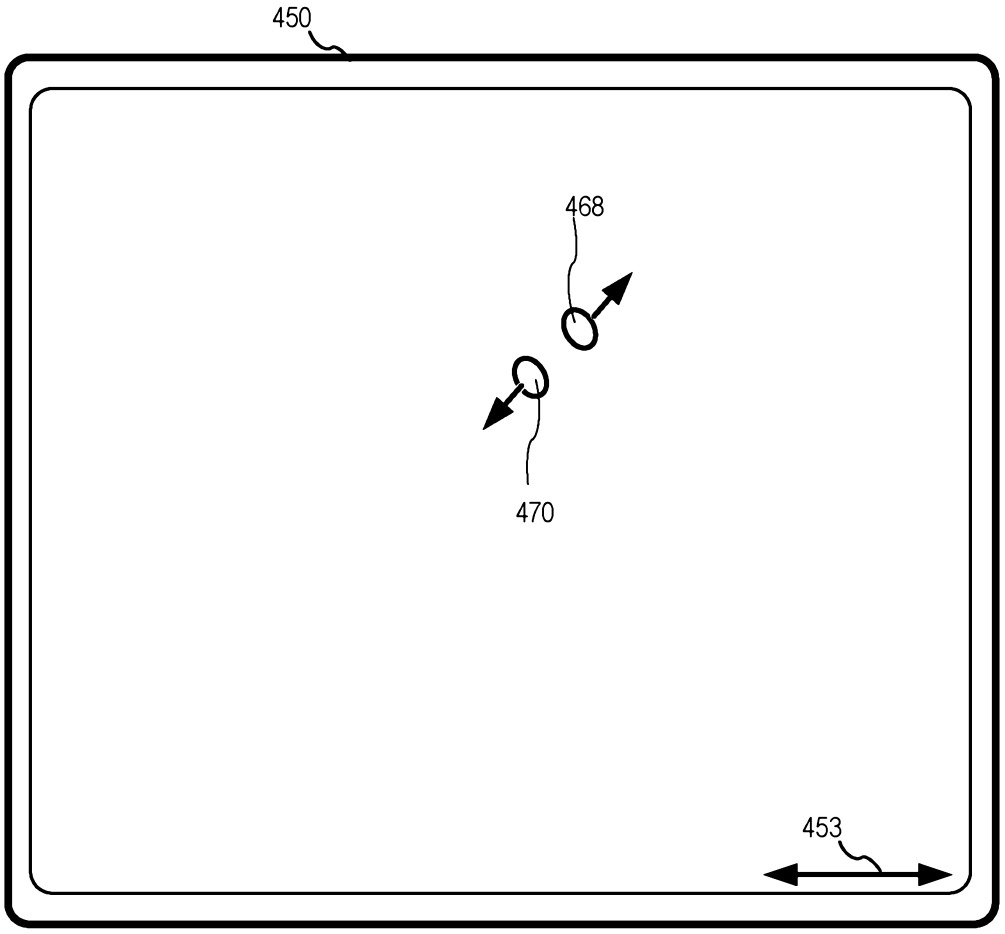

Portable Multifunction Device
100
Speaker 111
Optical Sensor 164
Proximity Sensor 166
400
208
208
206
405
402
Current Time 404
406
Messages 424
Tuesday
30
Calendar 426
Photos 428
Camera 430
Online Video 432
Stocks 434
280
Maps 436
73°
Weather 438
Clock 440
Workout Support 442
Notes 444
Settings 446
App Store
iTunes
Voice Memos
Utilities
414
4
6
410
Phone 416
Mail 418
Browser 420
iPod 422
408
Touch Screen 112
Microphone 113
Accelerometer(s) 168
204

FIG. 4A


450
468
470
453

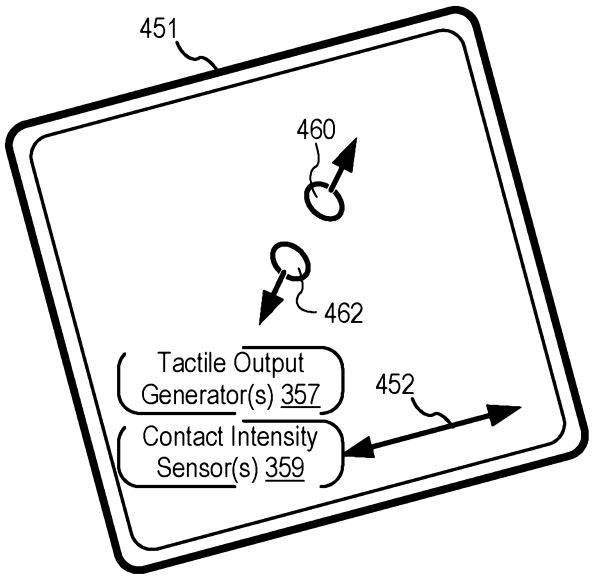

451
460
462
Tactile Output Generator(s) 357
452
Contact Intensity Sensor(s) 359

FIG. 4B

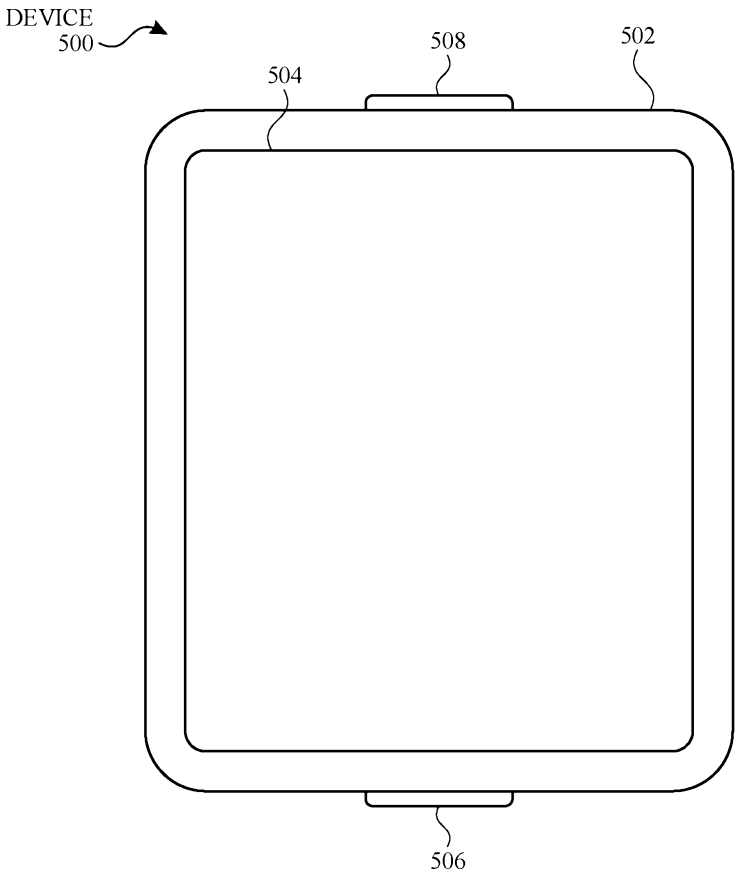


FIG. 5A

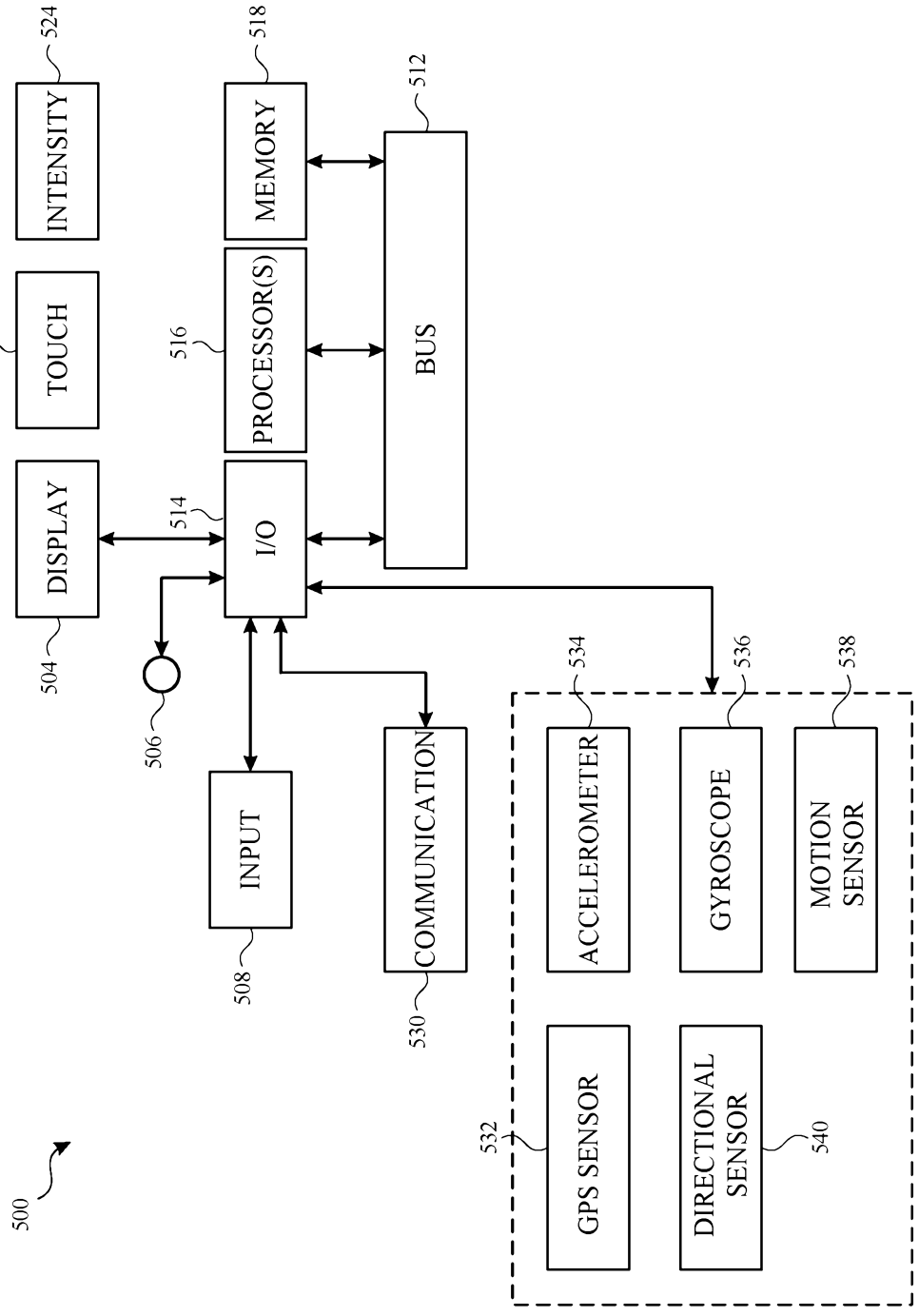

500
DISPLAY
504
TOUCH
522
INTENSITY
524
I/O
514
PROCESSOR(S)
516
MEMORY
518
BUS
512
506
INPUT
508
COMMUNICATION
530
ACCELEROMETER
534
GYROSCOPE
536
MOTION SENSOR
538
GPS SENSOR
532
DIRECTIONAL SENSOR
540

FIG. 5B

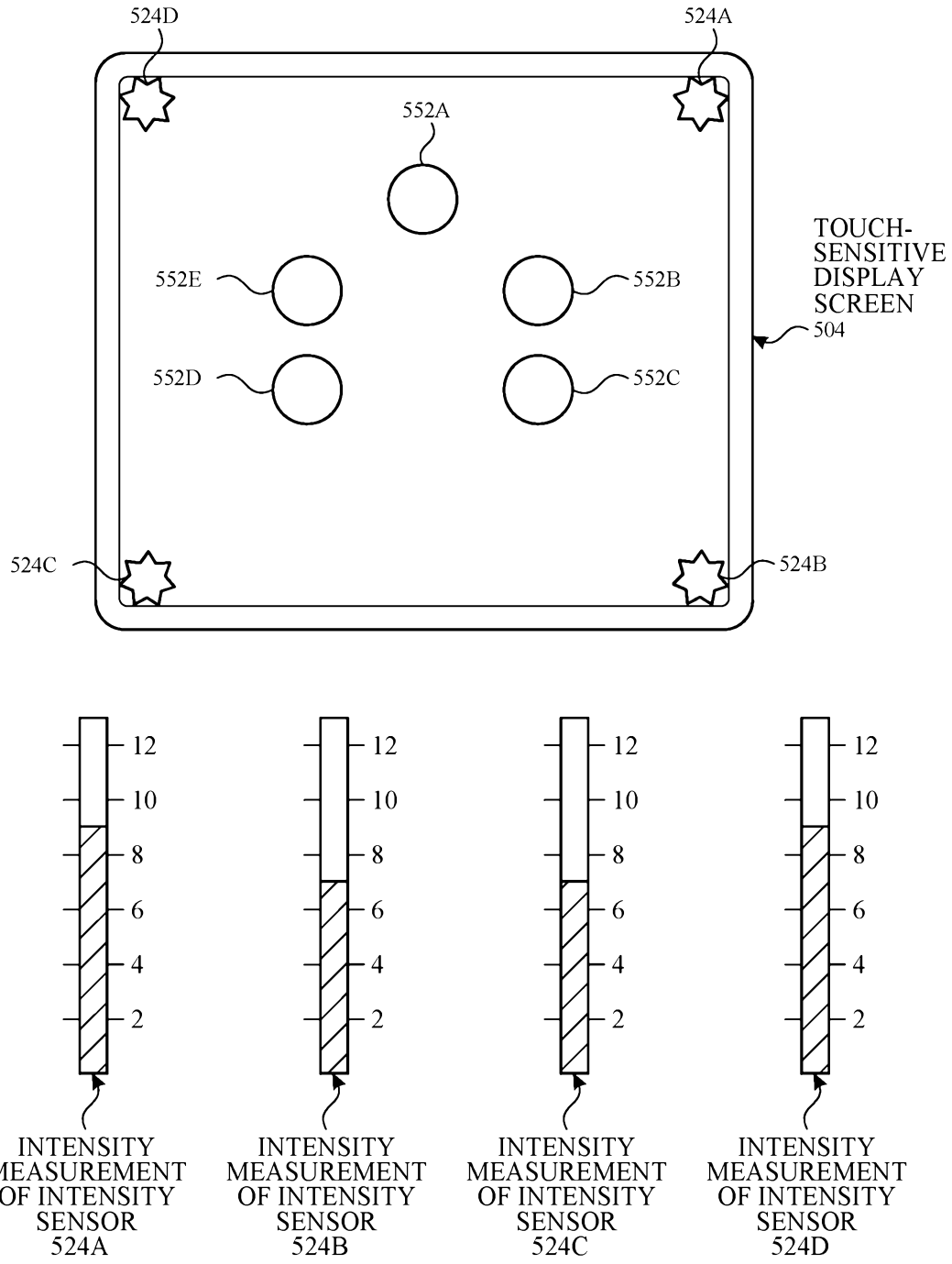

524D
524A
552A
552E
552B
552D
552C
TOUCH-SENSITIVE DISPLAY SCREEN
504
524C
524B
12
10
8
6
4
2
INTENSITY MEASUREMENT OF INTENSITY SENSOR 524A
INTENSITY MEASUREMENT OF INTENSITY SENSOR 524B
INTENSITY MEASUREMENT OF INTENSITY SENSOR 524C
INTENSITY MEASUREMENT OF INTENSITY SENSOR 524D

FIG. 5C

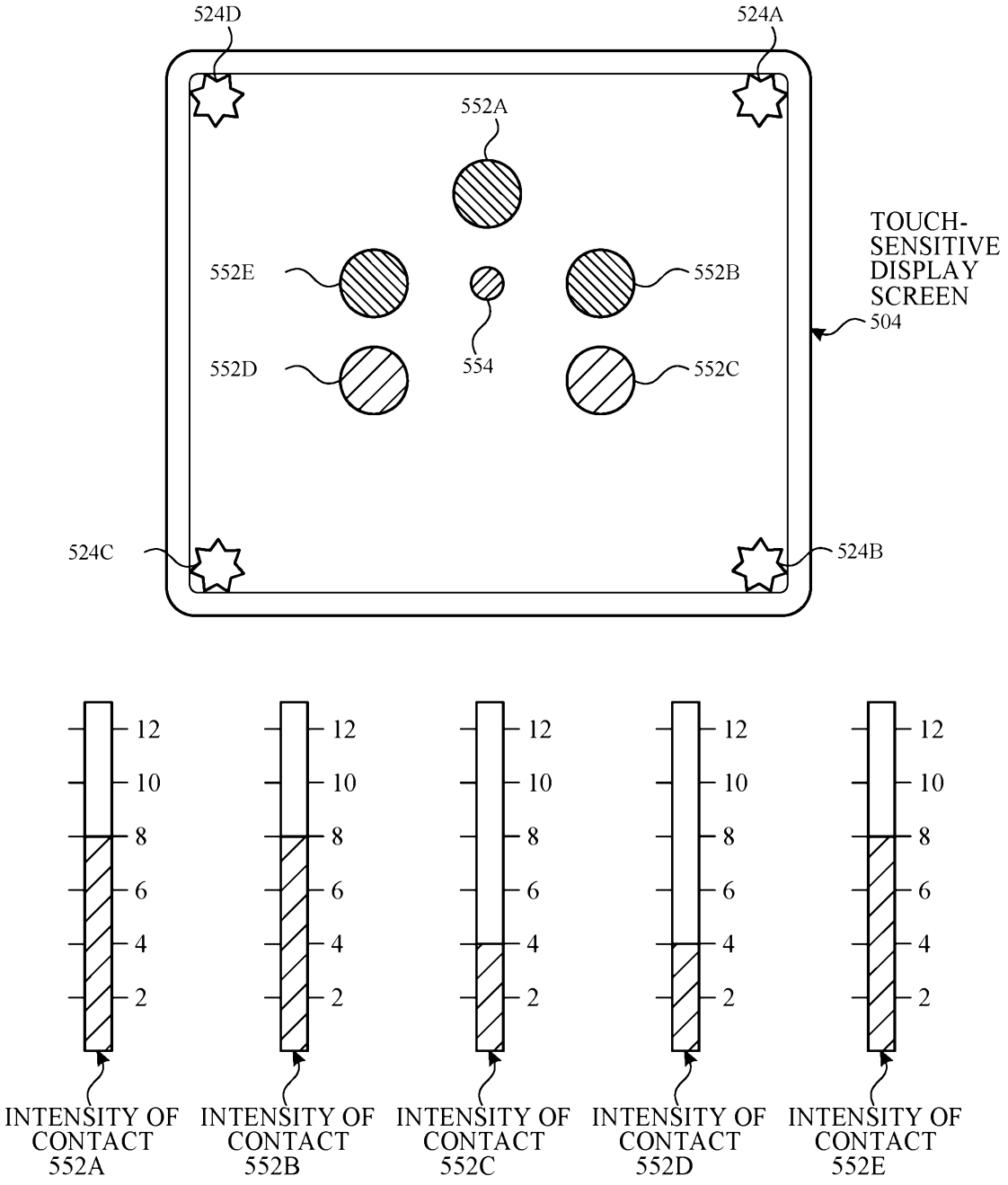

524D
524A
552A
552E
552B
554
552D
552C
TOUCH-SENSITIVE DISPLAY SCREEN
504
524C
524B
12
10
8
6
4
2
INTENSITY OF CONTACT 552A
INTENSITY OF CONTACT 552B
INTENSITY OF CONTACT 552C
INTENSITY OF CONTACT 552D
INTENSITY OF CONTACT 552E

FIG. 5D

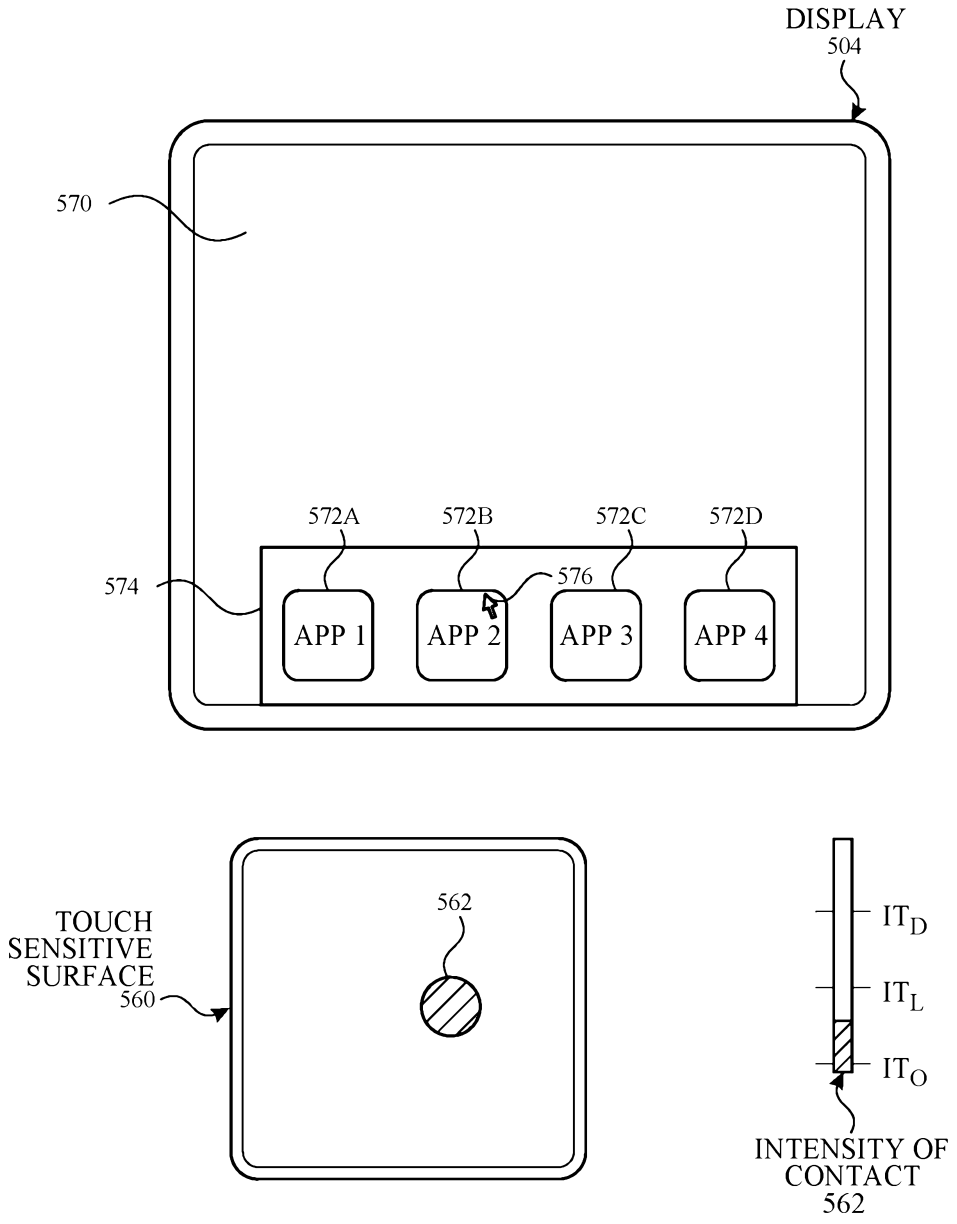

DISPLAY
504
570
572A
572B
572C
572D
574
576
APP 1
APP 2
APP 3
APP 4
TOUCH
SENSITIVE
SURFACE
560
562
IT$_D$
IT$_L$
IT$_O$
INTENSITY OF
CONTACT
562

FIG. 5E

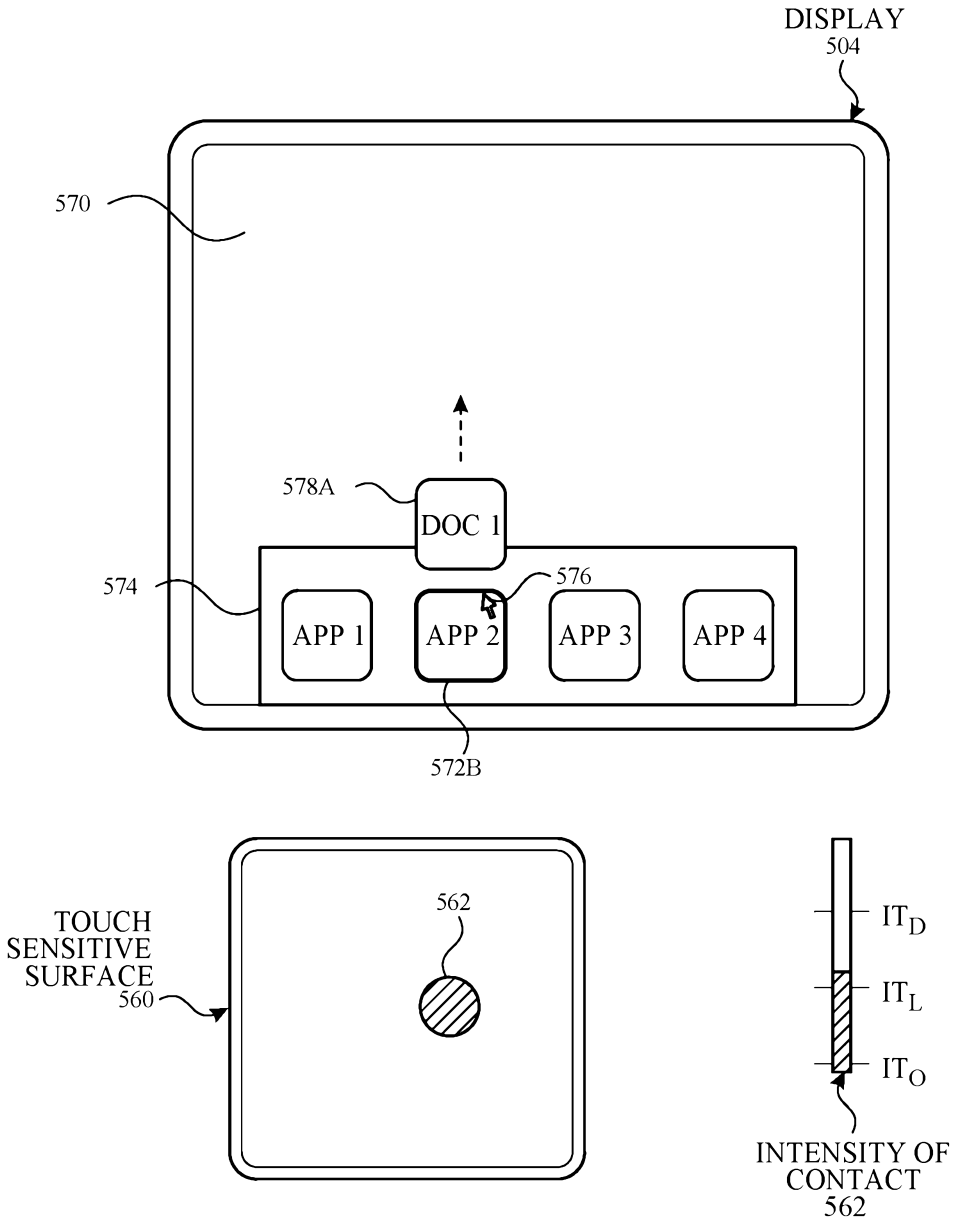

DISPLAY
504
570
578A
DOC 1
574
576
APP 1
APP 2
APP 3
APP 4
572B
TOUCH
SENSITIVE
SURFACE
560
562
IT$_D$
IT$_L$
IT$_O$
INTENSITY OF
CONTACT
562

FIG. 5F

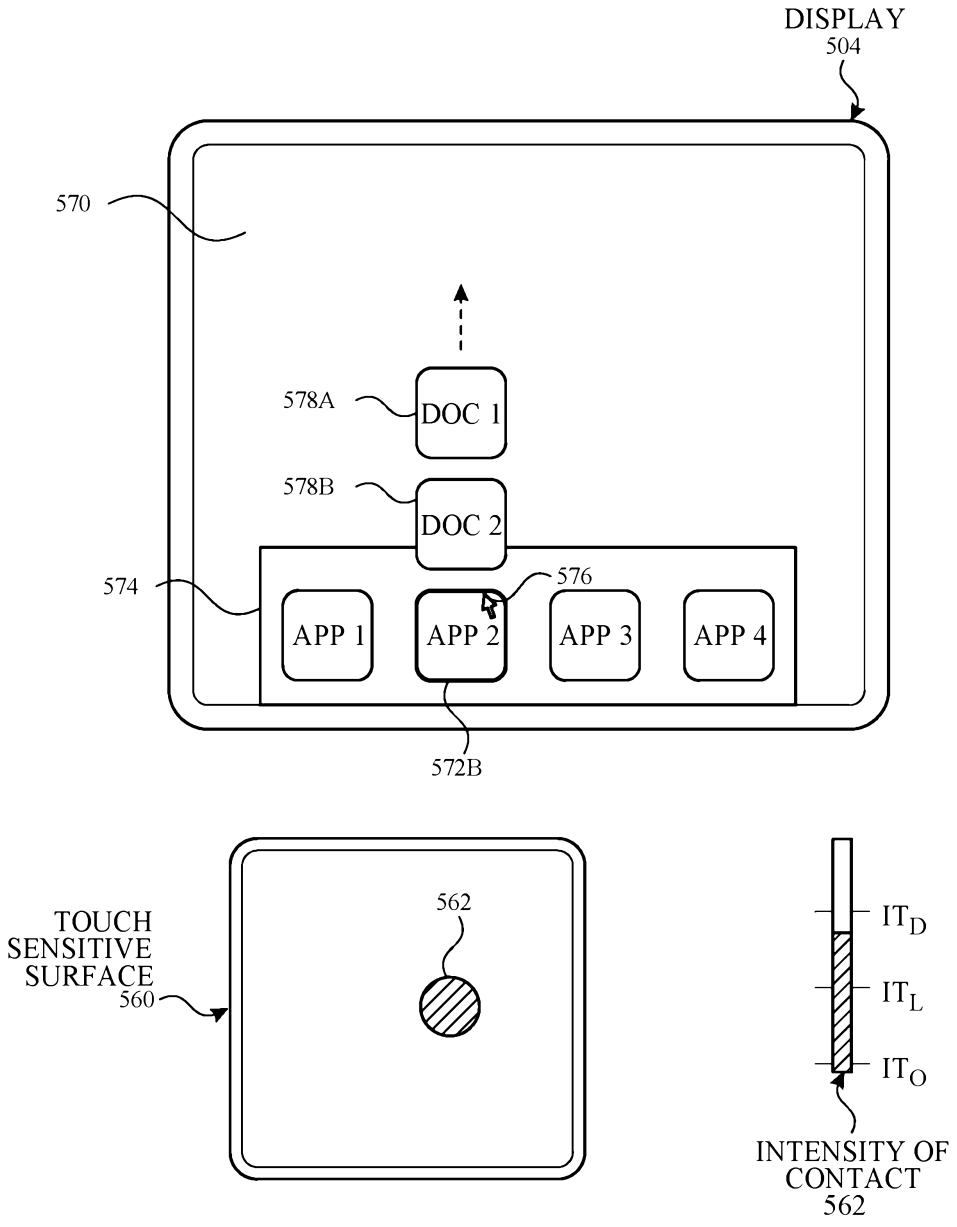

DISPLAY
504
570
578A
DOC 1
578B
DOC 2
574
576
APP 1
APP 2
APP 3
APP 4
572B
TOUCH
SENSITIVE
SURFACE
560
562
IT$_D$
IT$_L$
IT$_O$
INTENSITY OF
CONTACT
562

FIG. 5G

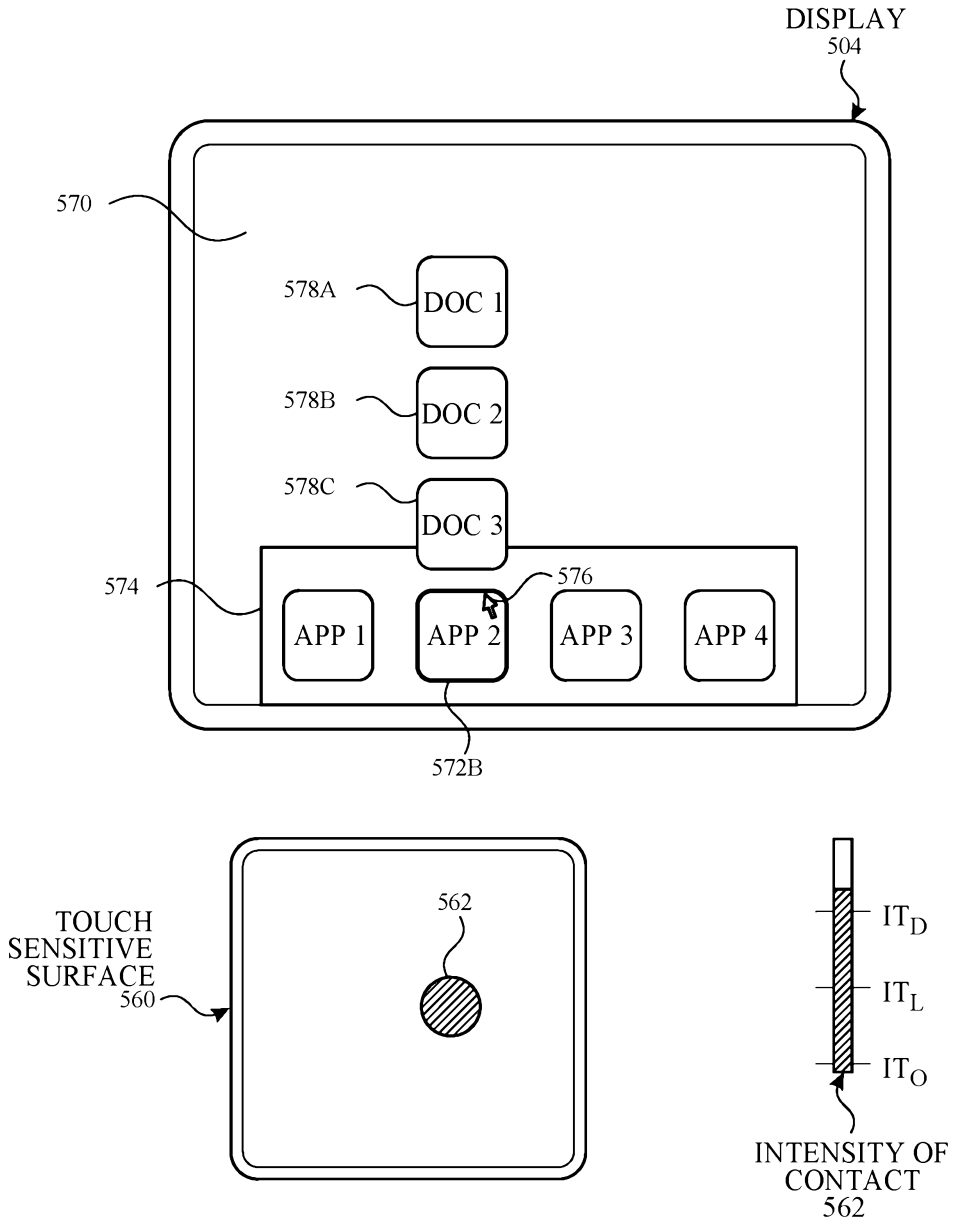

DISPLAY
504
570
578A
DOC 1
578B
DOC 2
578C
DOC 3
574
576
APP 1
APP 2
APP 3
APP 4
572B
TOUCH
SENSITIVE
SURFACE
560
562
IT_D
IT_L
IT_O
INTENSITY OF
CONTACT
562

FIG. 5H

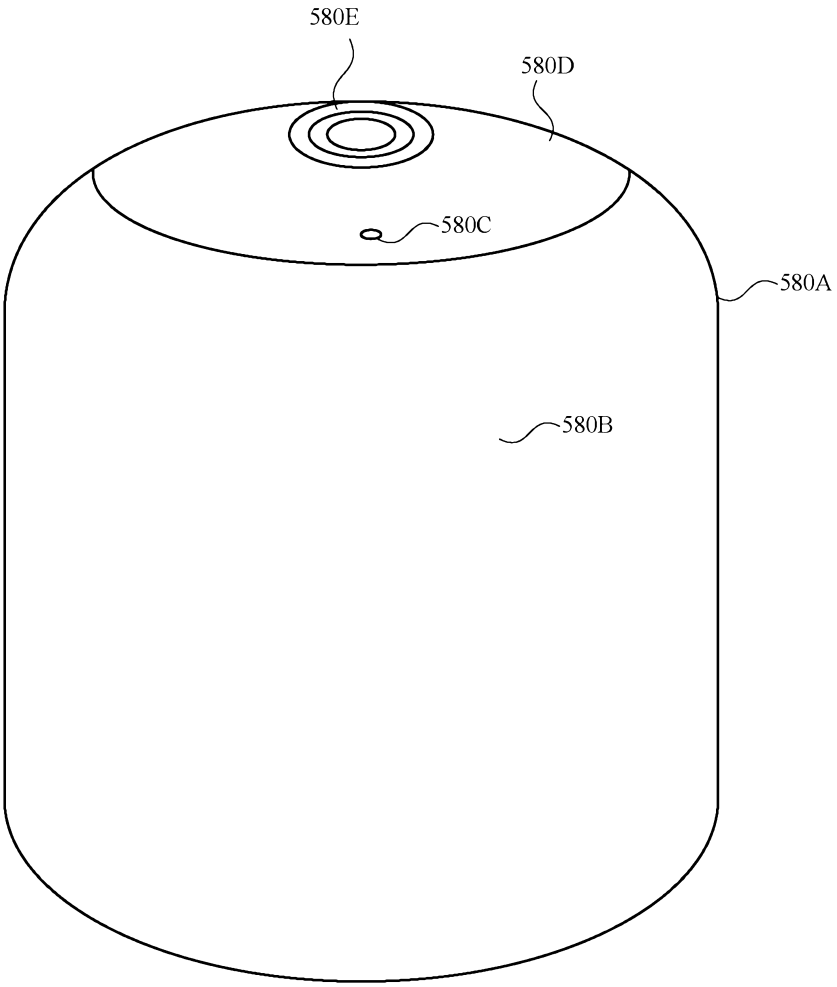


FIG. 5I

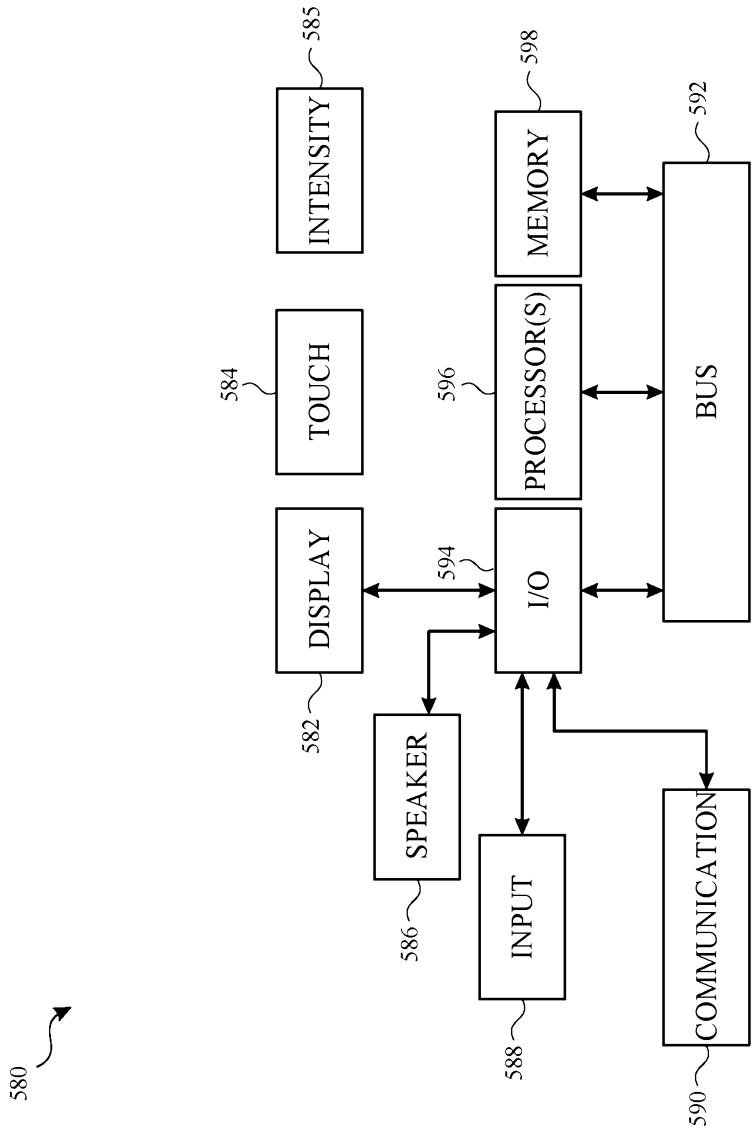

580
582
DISPLAY
584
TOUCH
585
INTENSITY
586
SPEAKER
588
INPUT
590
COMMUNICATION
592
BUS
594
I/O
596
PROCESSOR(S)
598
MEMORY

FIG. 5J

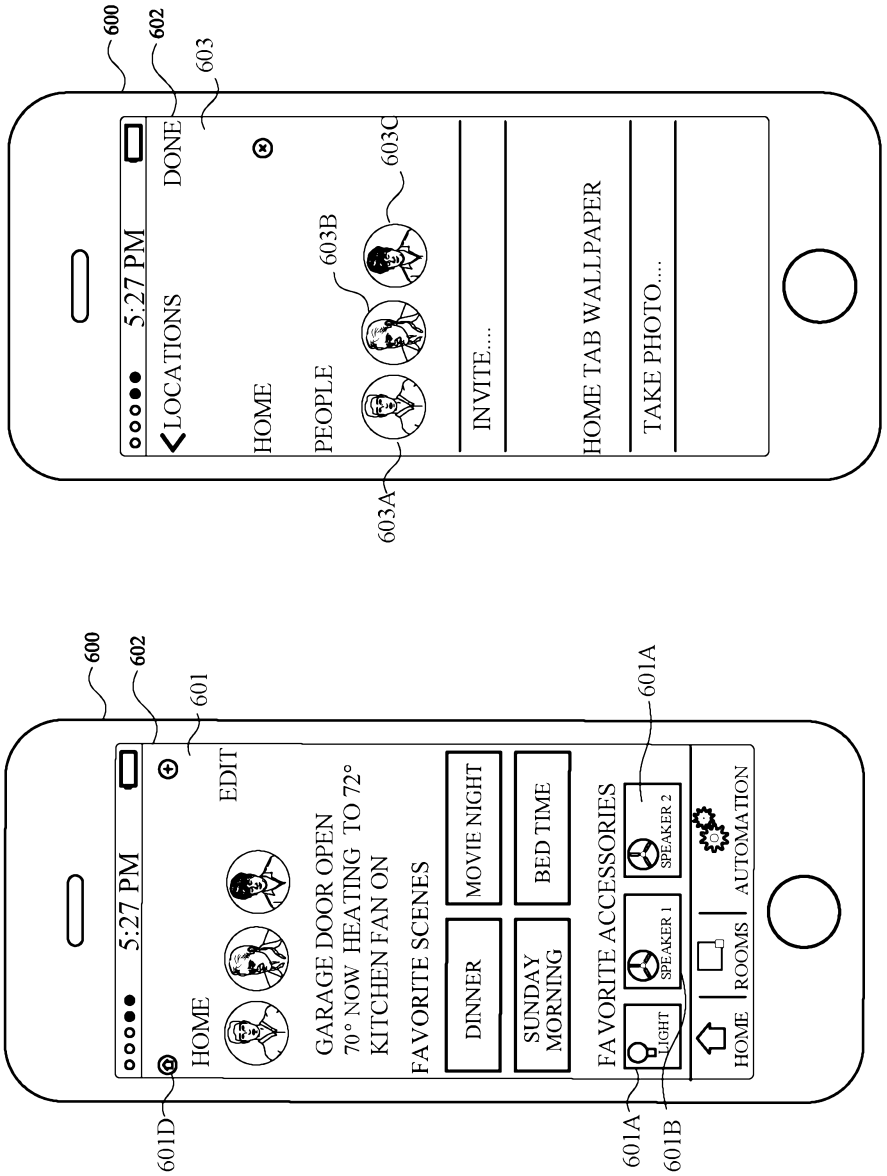

600
602
603
5:27 PM
DONE
LOCATIONS
HOME
PEOPLE
603B
603C
INVITE....
HOME TAB WALLPAPER
TAKE PHOTO....
603A
600
602
601
5:27 PM
EDIT
HOME
GARAGE DOOR OPEN
70° NOW HEATING TO 72°
KITCHEN FAN ON
FAVORITE SCENES
DINNER
MOVIE NIGHT
SUNDAY MORNING
BED TIME
FAVORITE ACCESSORIES
601A
LIGHT
SPEAKER 1
SPEAKER 2
HOME
ROOMS
AUTOMATION
601D
601A
601B

FIG. 6A

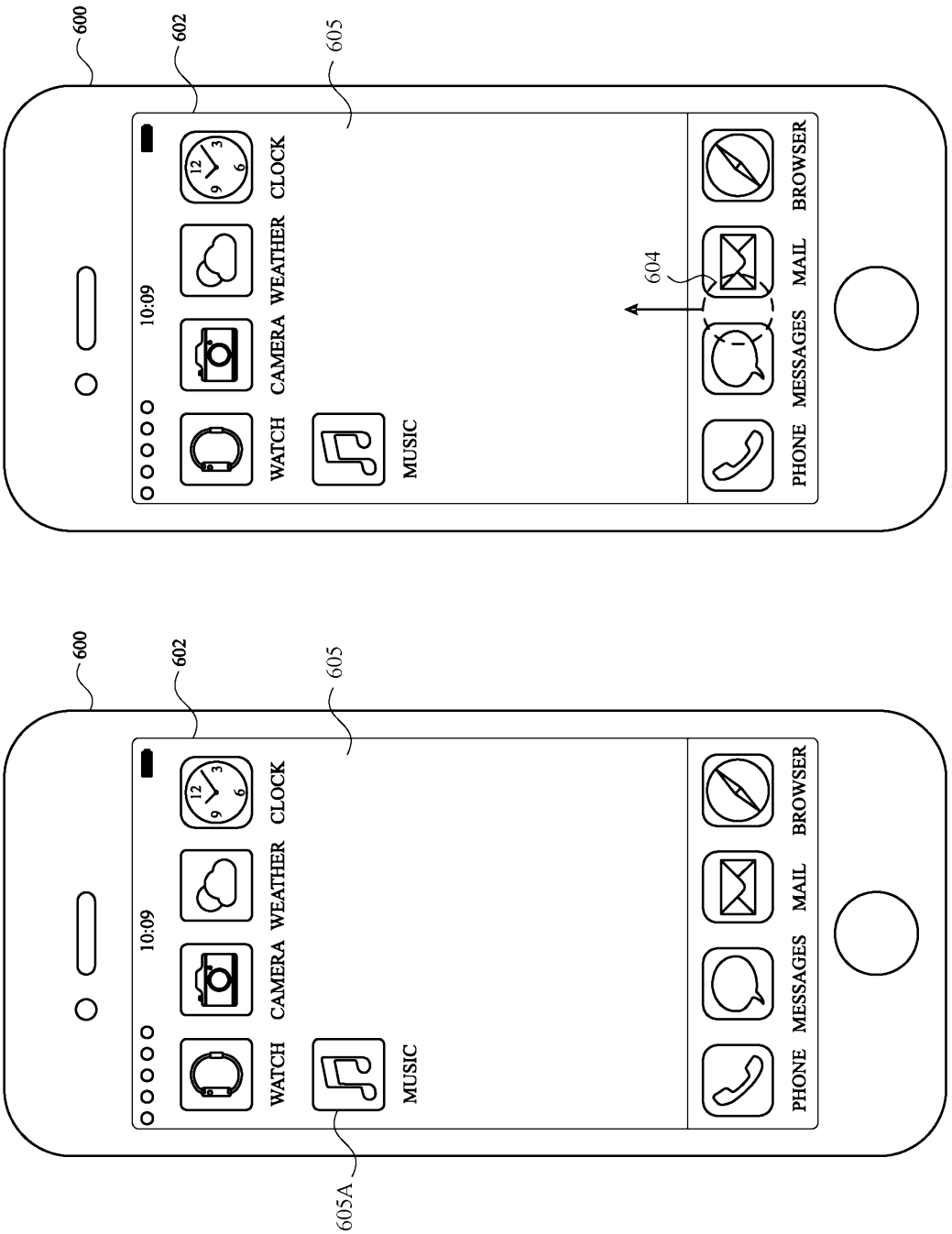

600
602
605
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
604
PHONE
MESSAGES
MAIL
BROWSER
600
602
605
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
605A
PHONE
MESSAGES
MAIL
BROWSER

FIG. 6B

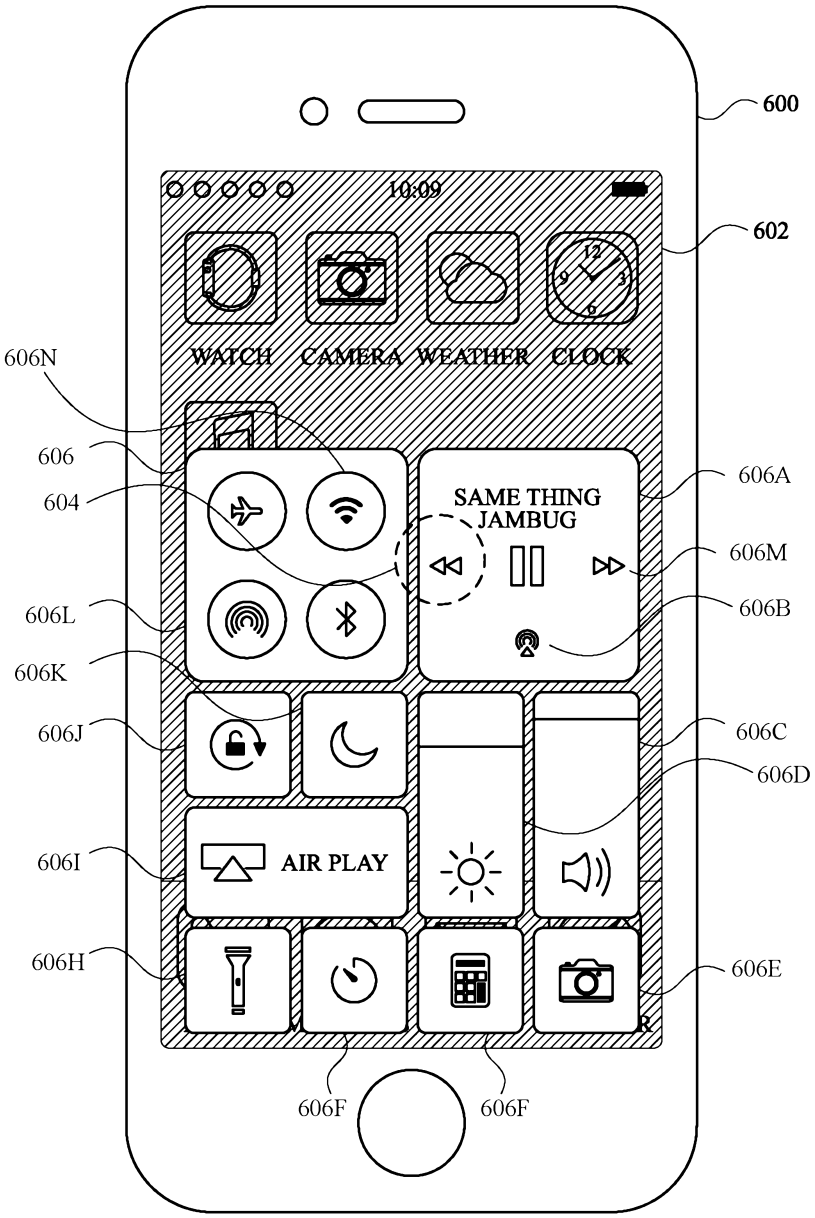

600
10:09
602
WATCH
CAMERA
WEATHER
CLOCK
606N
606
604
606A
SAME THING
JAMBUG
606M
606L
606B
606K
606J
606C
606D
AIR PLAY
606I
606H
606E
606F
606F

FIG. 6C

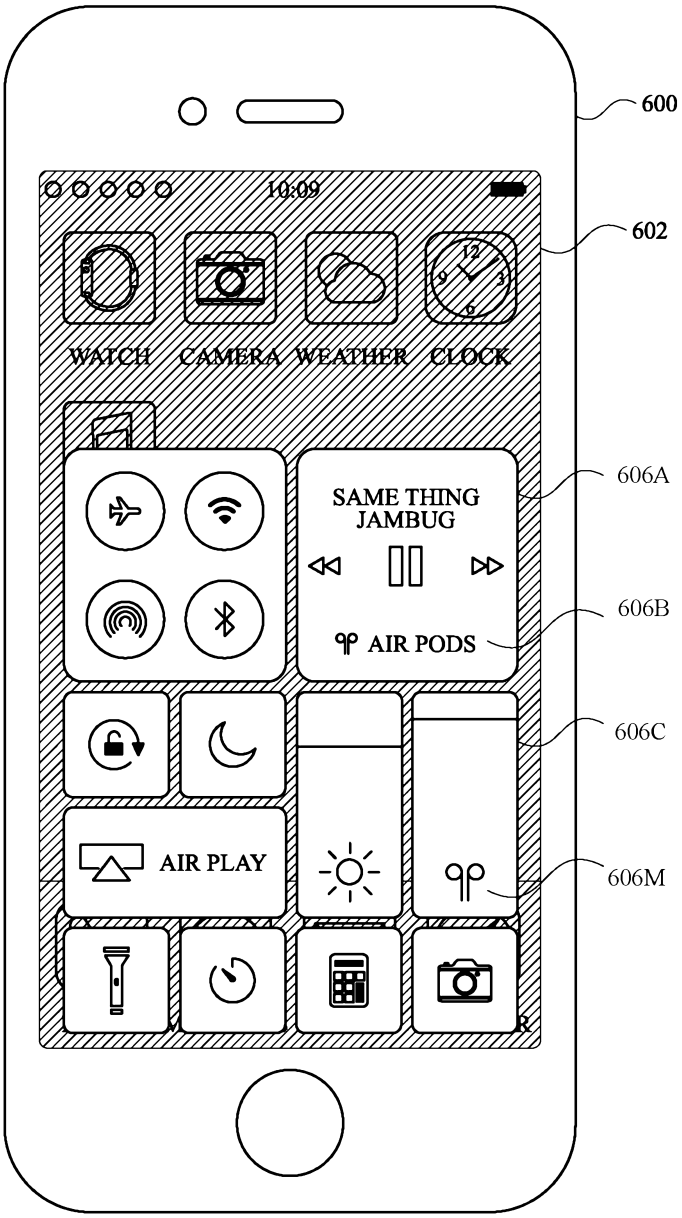

600
10:09
602
WATCH
CAMERA
WEATHER
CLOCK
606A
SAME THING
JAMBUG
606B
AIR PODS
606C
AIR PLAY
606M

FIG. 6D

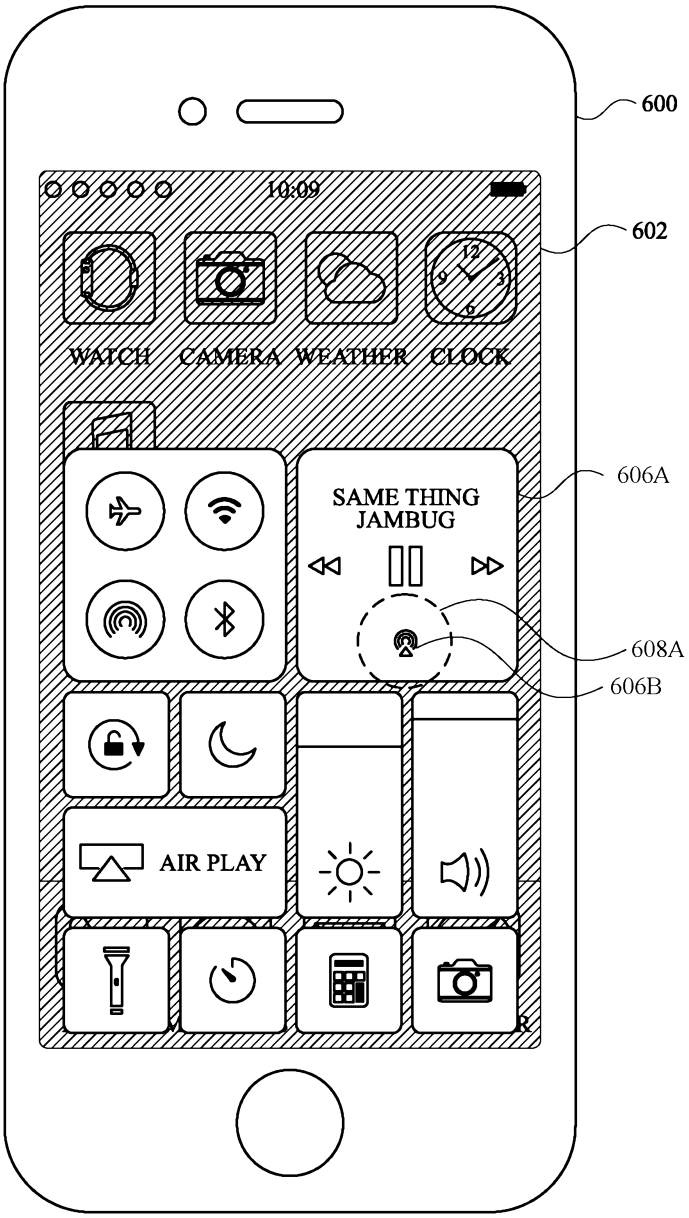

600
10:09
602
WATCH
CAMERA
WEATHER
CLOCK
606A
SAME THING
JAMBUG
608A
606B
AIR PLAY

FIG. 6E

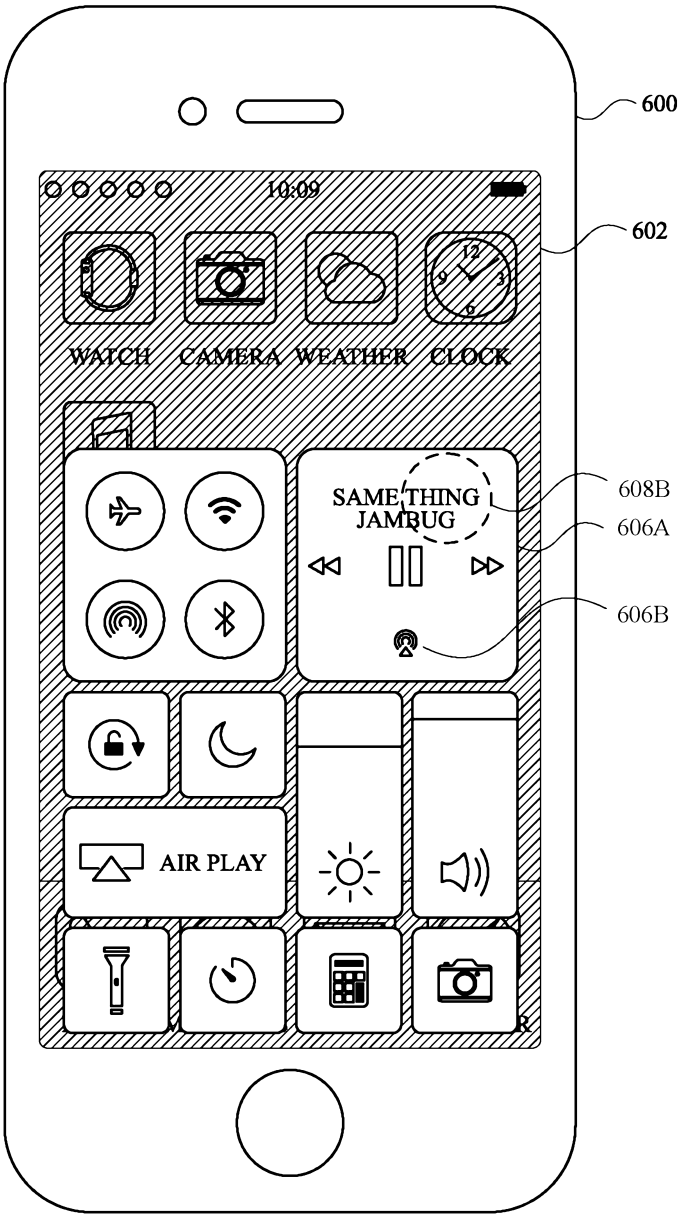

600
10:09
602
WATCH
CAMERA
WEATHER
CLOCK
SAME THING
JAMBUG
608B
606A
606B
AIR PLAY

FIG. 6F

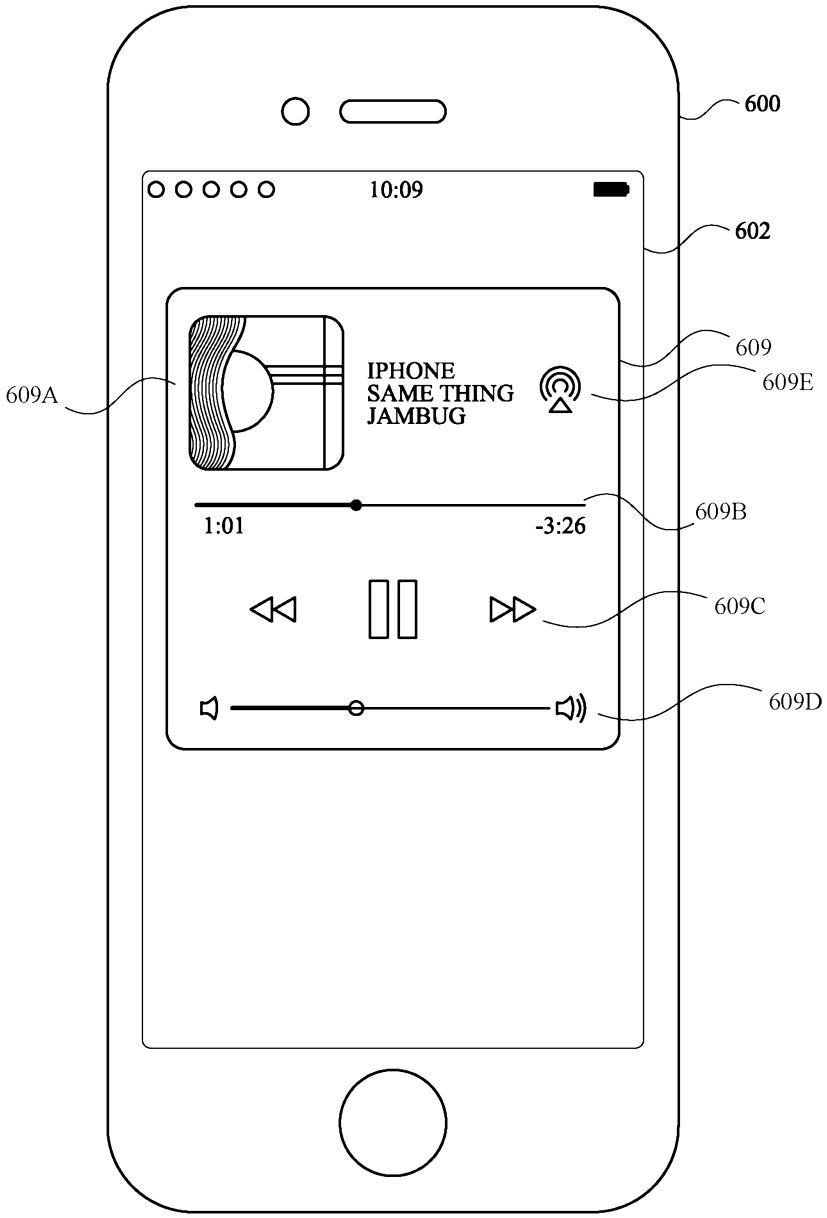

600
10:09
602
609
609A
IPHONE
SAME THING
JAMBUG
609E
609B
1:01
-3:26
609C
609D

FIG. 6G

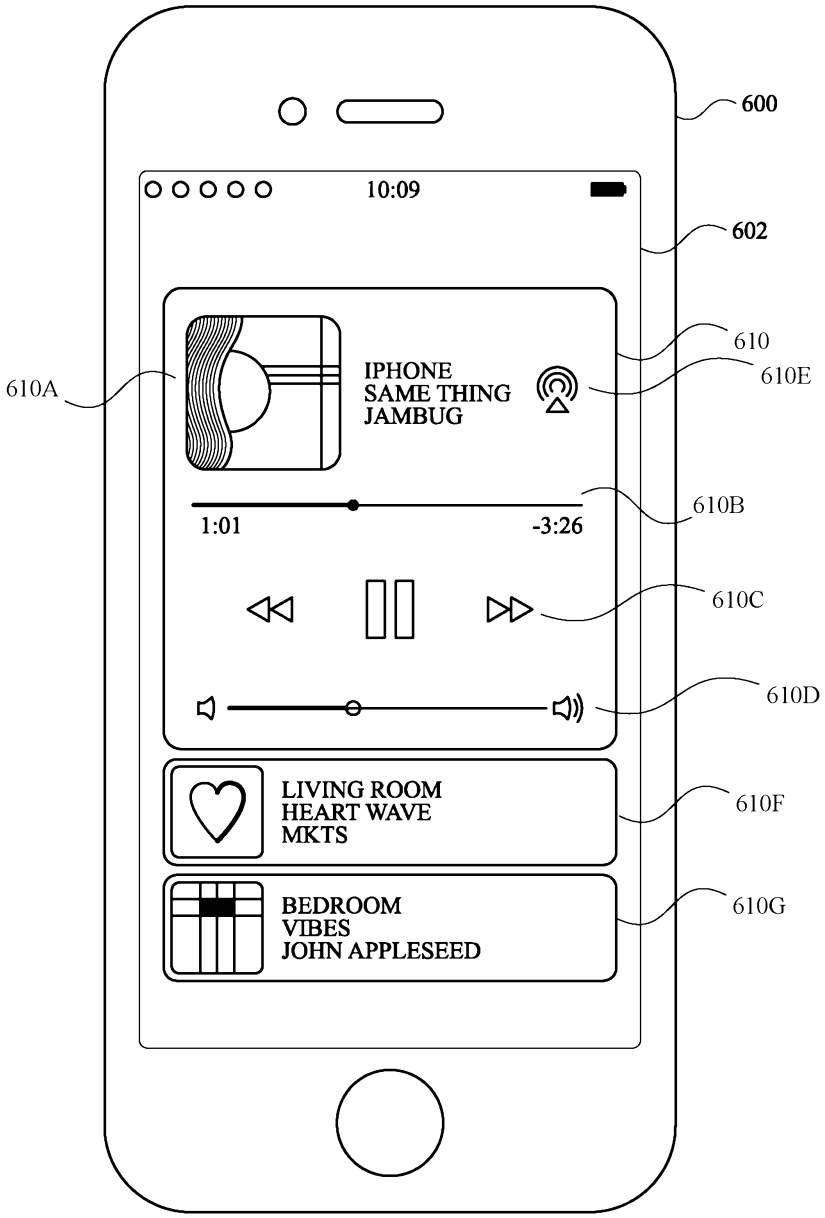

600
10:09
602
610
610A
IPHONE
SAME THING
JAMBUG
610E
610B
1:01
-3:26
610C
610D
LIVING ROOM
HEART WAVE
MKTS
610F
BEDROOM
VIBES
JOHN APPLESEED
610G

FIG. 6H

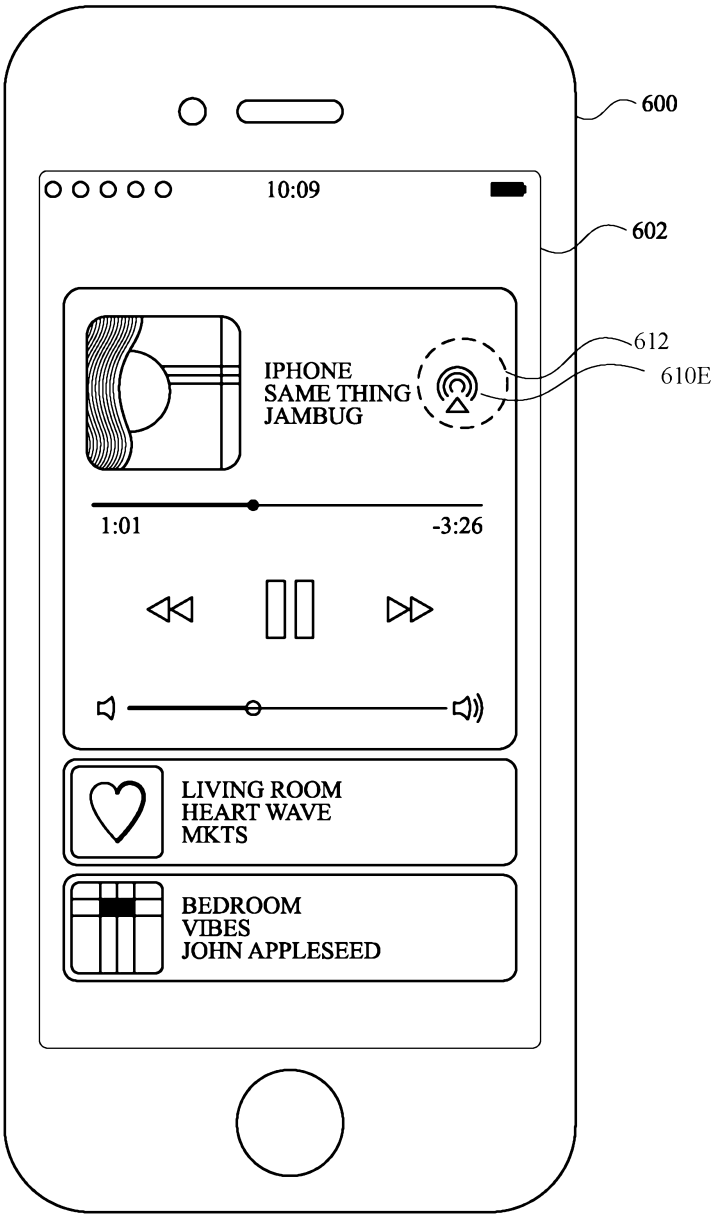

600
10:09
602
612
610E
IPHONE
SAME THING
JAMBUG
1:01
-3:26
LIVING ROOM
HEART WAVE
MKTS
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6I

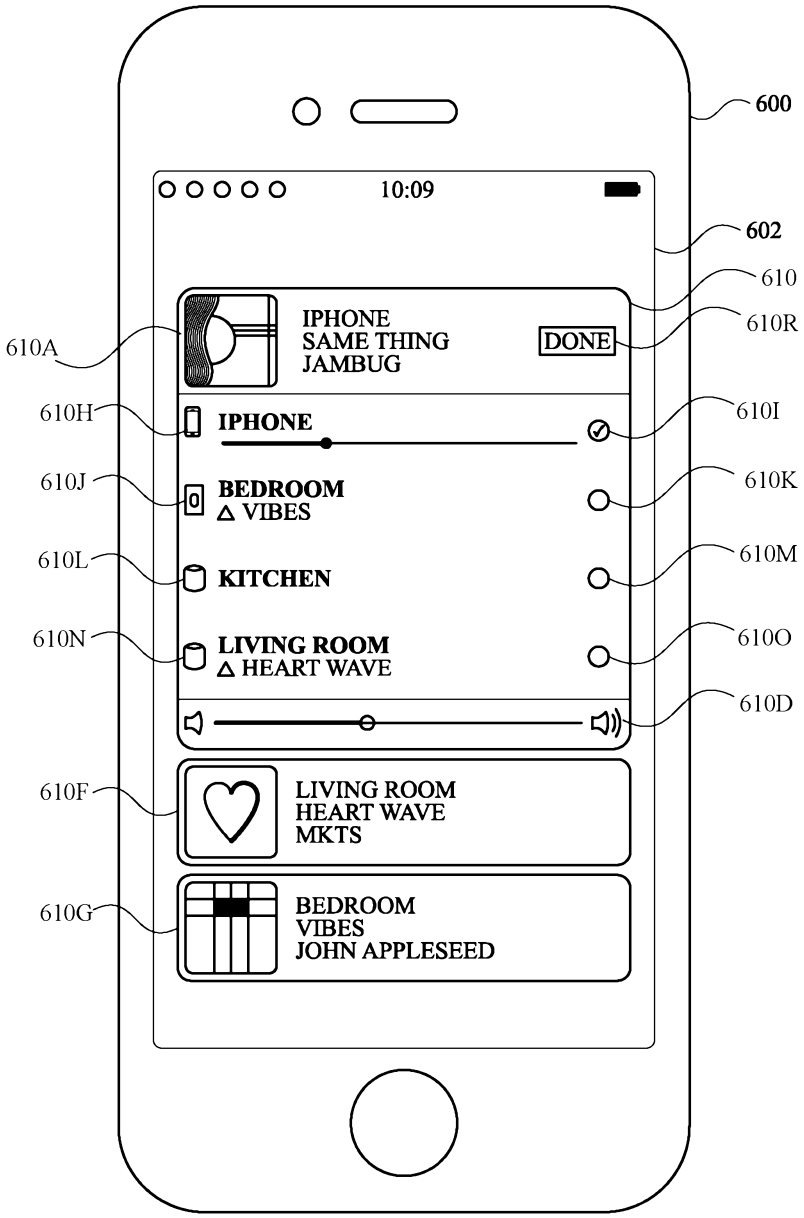

600
10:09
602
610
610R
610A
IPHONE
SAME THING
JAMBUG
DONE
610H
IPHONE
610I
610J
BEDROOM
△ VIBES
610K
610L
KITCHEN
610M
610N
LIVING ROOM
△ HEART WAVE
610O
610D
610F
LIVING ROOM
HEART WAVE
MKTS
610G
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6J

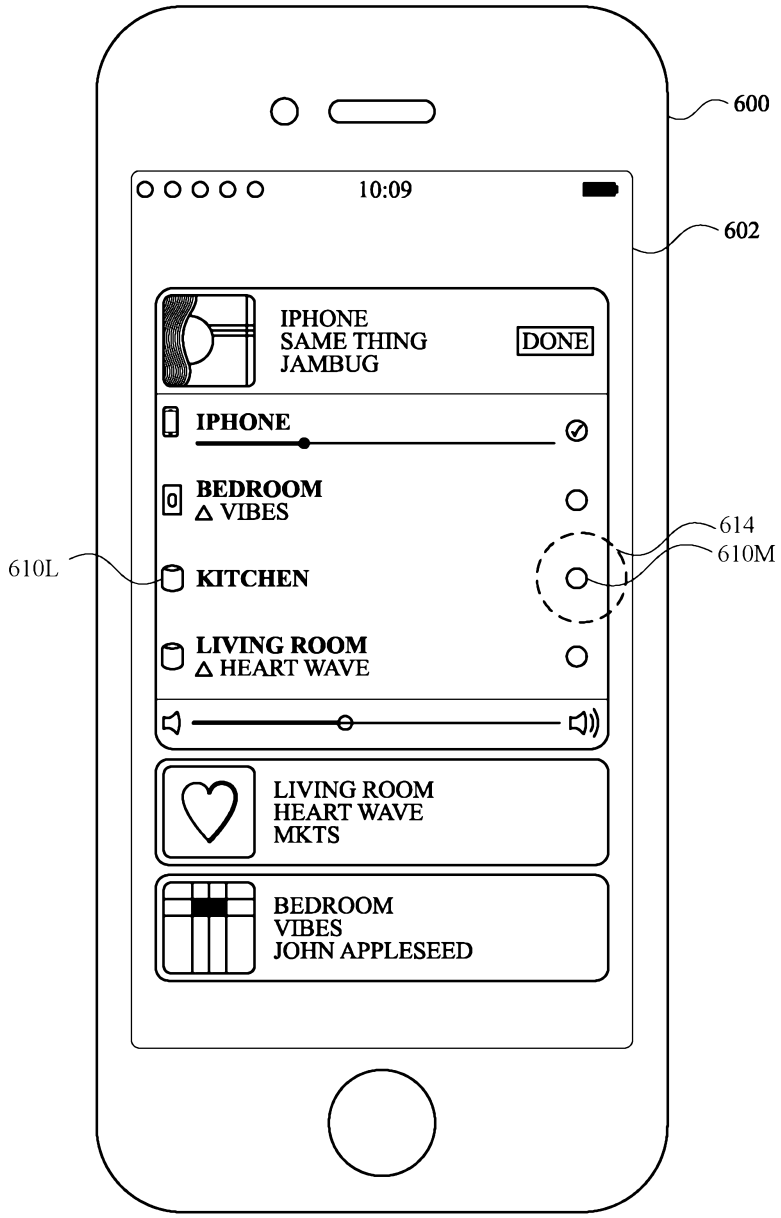

600
10:09
602
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
BEDROOM
VIBES
610L
KITCHEN
614
610M
LIVING ROOM
HEART WAVE
LIVING ROOM
HEART WAVE
MKTS
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6K

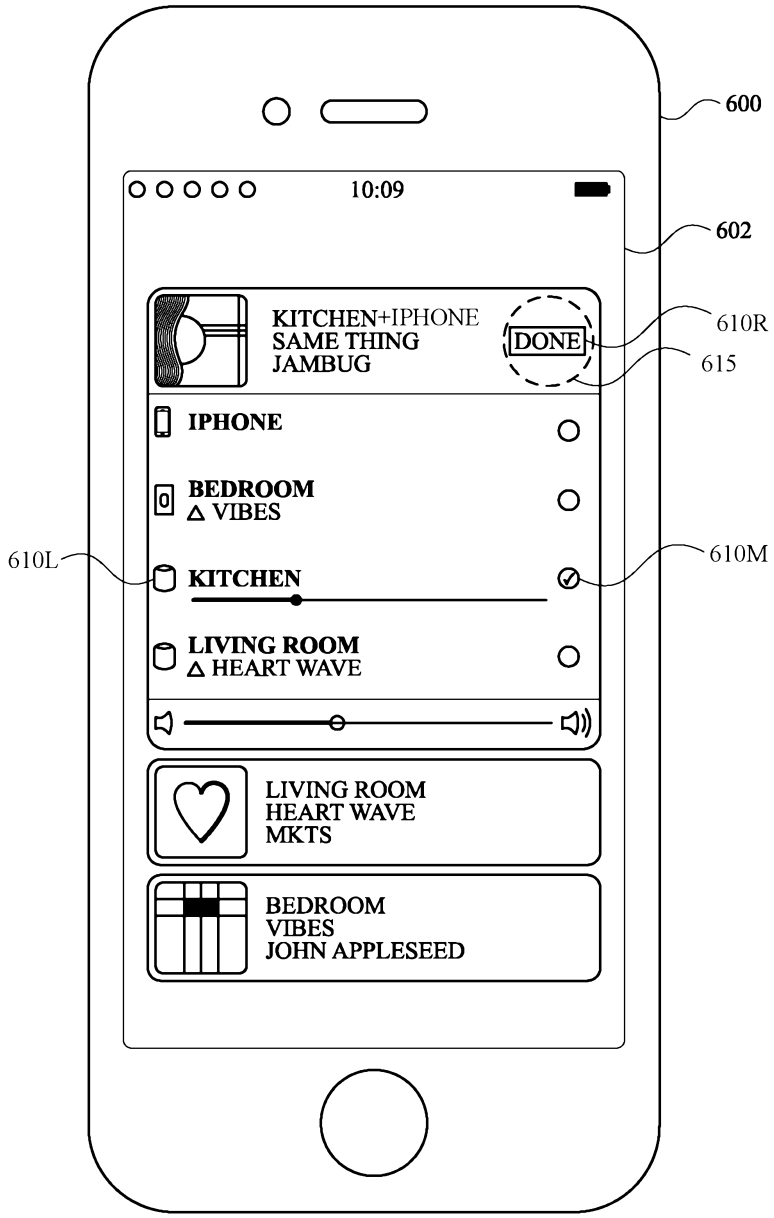

600
10:09
602
KITCHEN+IPHONE
SAME THING
JAMBUG
DONE
610R
615
IPHONE
BEDROOM
△ VIBES
610L
KITCHEN
610M
LIVING ROOM
△ HEART WAVE
LIVING ROOM
HEART WAVE
MKTS
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6L

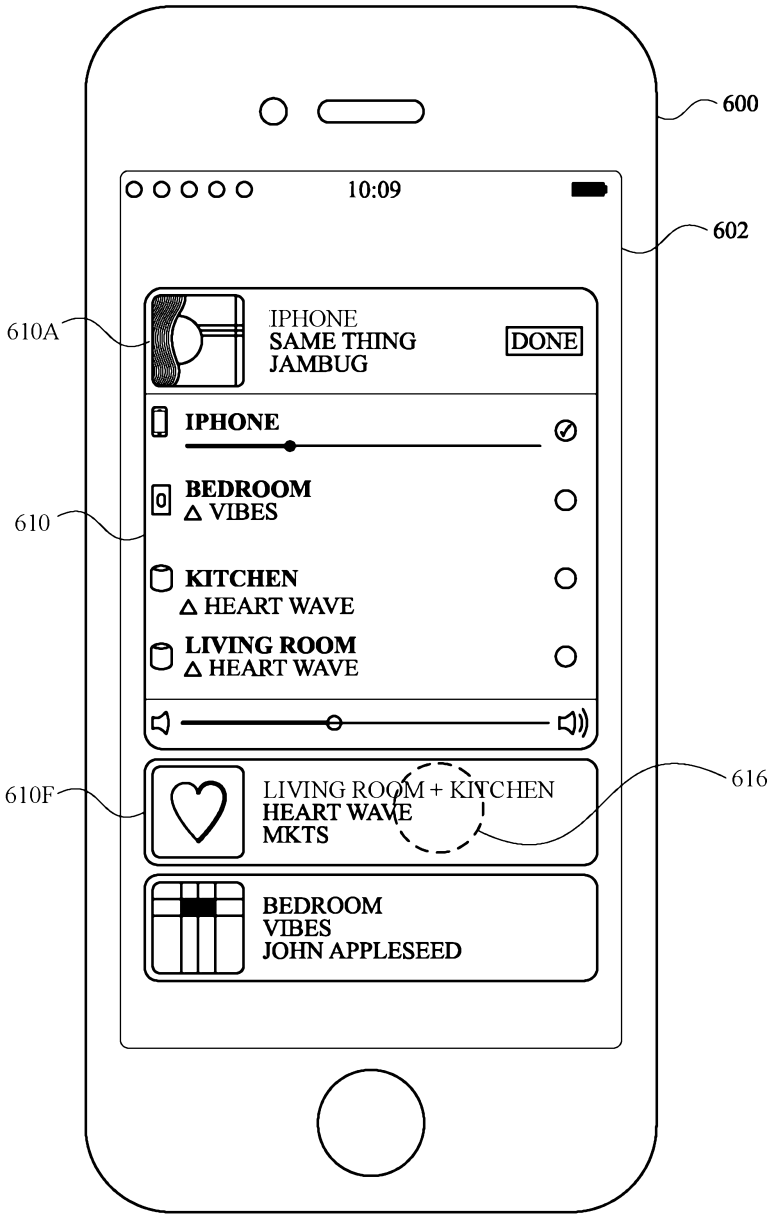

600
10:09
602
610A
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
610
BEDROOM
△ VIBES
KITCHEN
△ HEART WAVE
LIVING ROOM
△ HEART WAVE
610F
LIVING ROOM + KITCHEN
HEART WAVE
MKTS
616
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6M

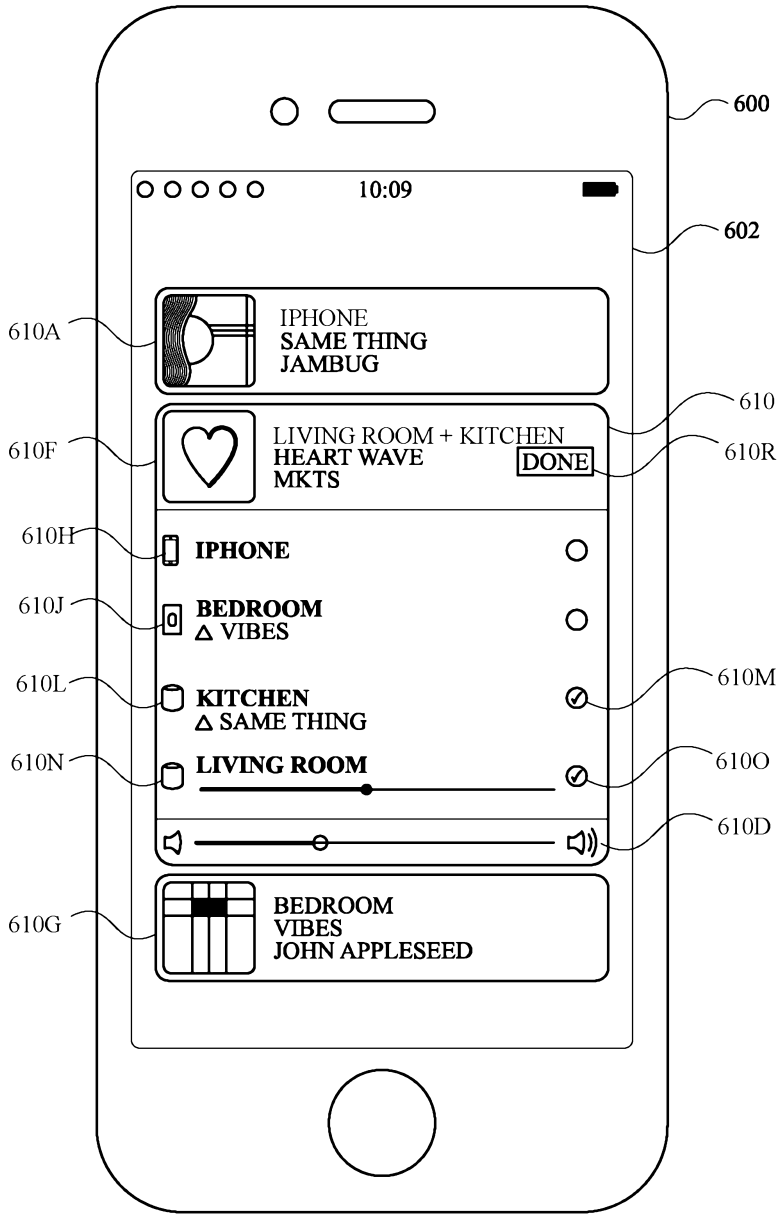

600
10:09
602
610A
IPHONE
SAME THING
JAMBUG
610
610F
LIVING ROOM + KITCHEN
HEART WAVE
MKTS
DONE
610R
610H
IPHONE
610J
BEDROOM
△ VIBES
610L
610M
KITCHEN
△ SAME THING
610N
LIVING ROOM
610O
610D
610G
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6N

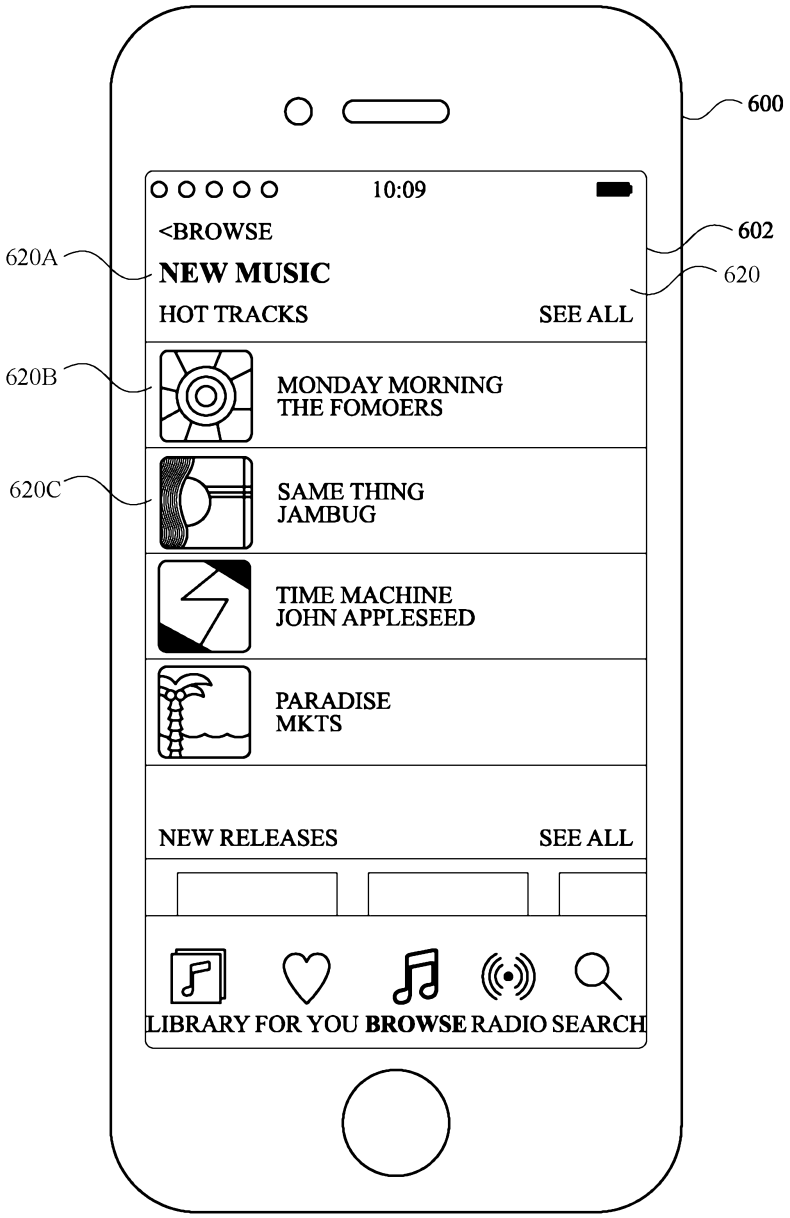

600
10:09
<BROWSE
602
620A
NEW MUSIC
620
HOT TRACKS
SEE ALL
620B
MONDAY MORNING
THE FOMOERS
620C
SAME THING
JAMBUG
TIME MACHINE
JOHN APPLESEED
PARADISE
MKTS
NEW RELEASES
SEE ALL
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 6O

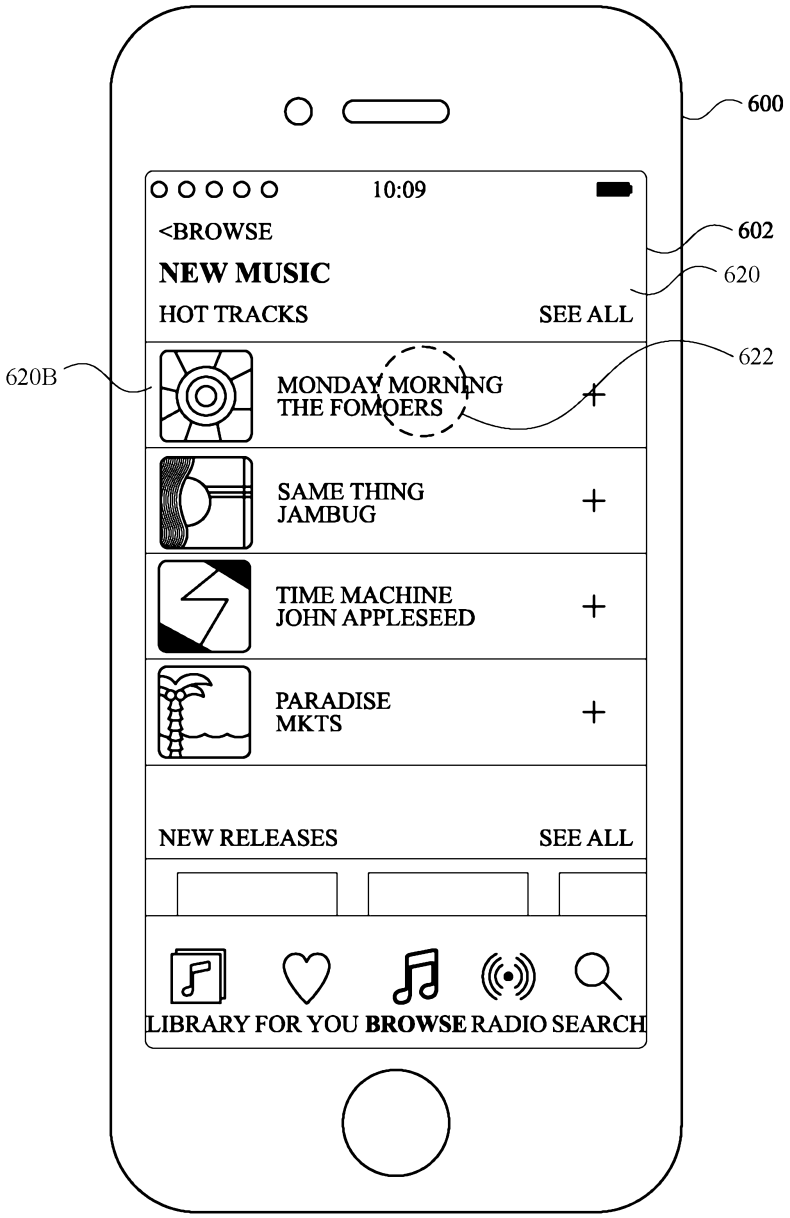

600
10:09
<BROWSE
602
NEW MUSIC
620
HOT TRACKS
SEE ALL
620B
622
MONDAY MORNING
THE FOMOERS
SAME THING
JAMBUG
TIME MACHINE
JOHN APPLESEED
PARADISE
MKTS
NEW RELEASES
SEE ALL
LIBRARY FOR YOU
BROWSE
RADIO
SEARCH

FIG. 6P

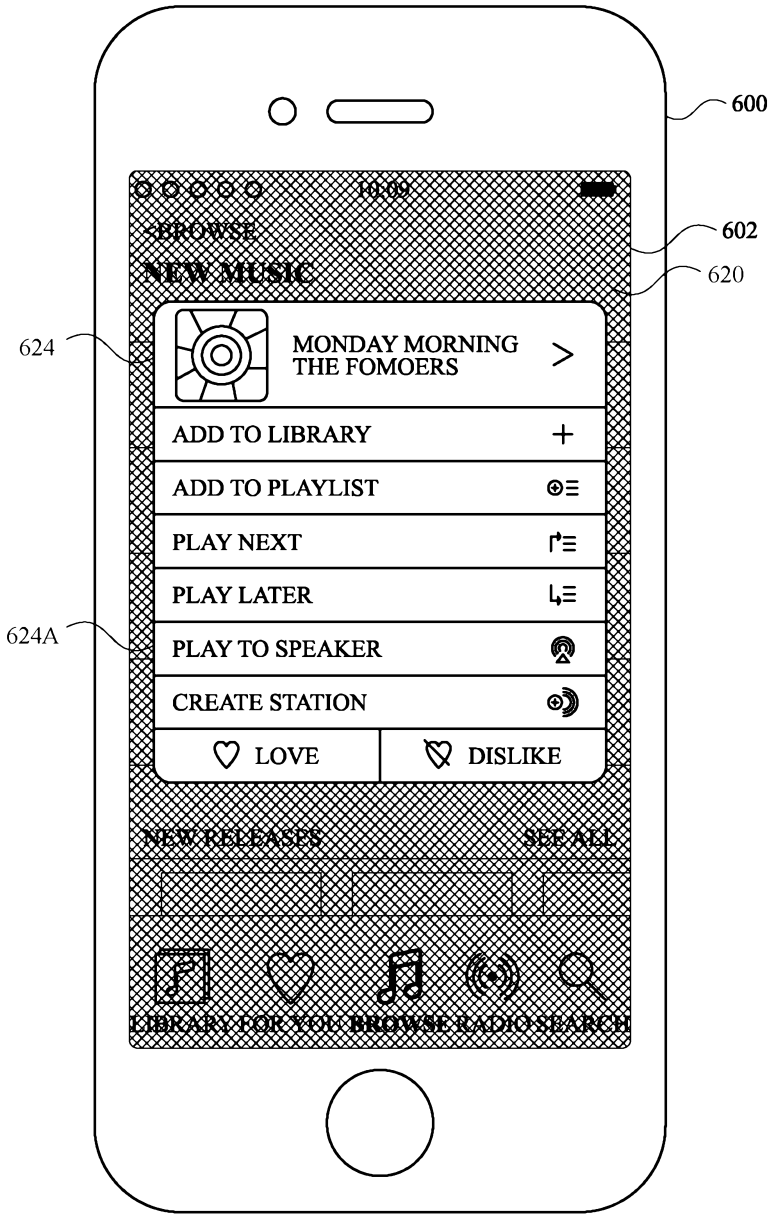

600
602
620
624
624A
10:09
<BROWSE
NEW MUSIC
MONDAY MORNING
THE FOMOERS
ADD TO LIBRARY
ADD TO PLAYLIST
PLAY NEXT
PLAY LATER
PLAY TO SPEAKER
CREATE STATION
LOVE
DISLIKE
NEW RELEASES
SEE ALL
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 6Q

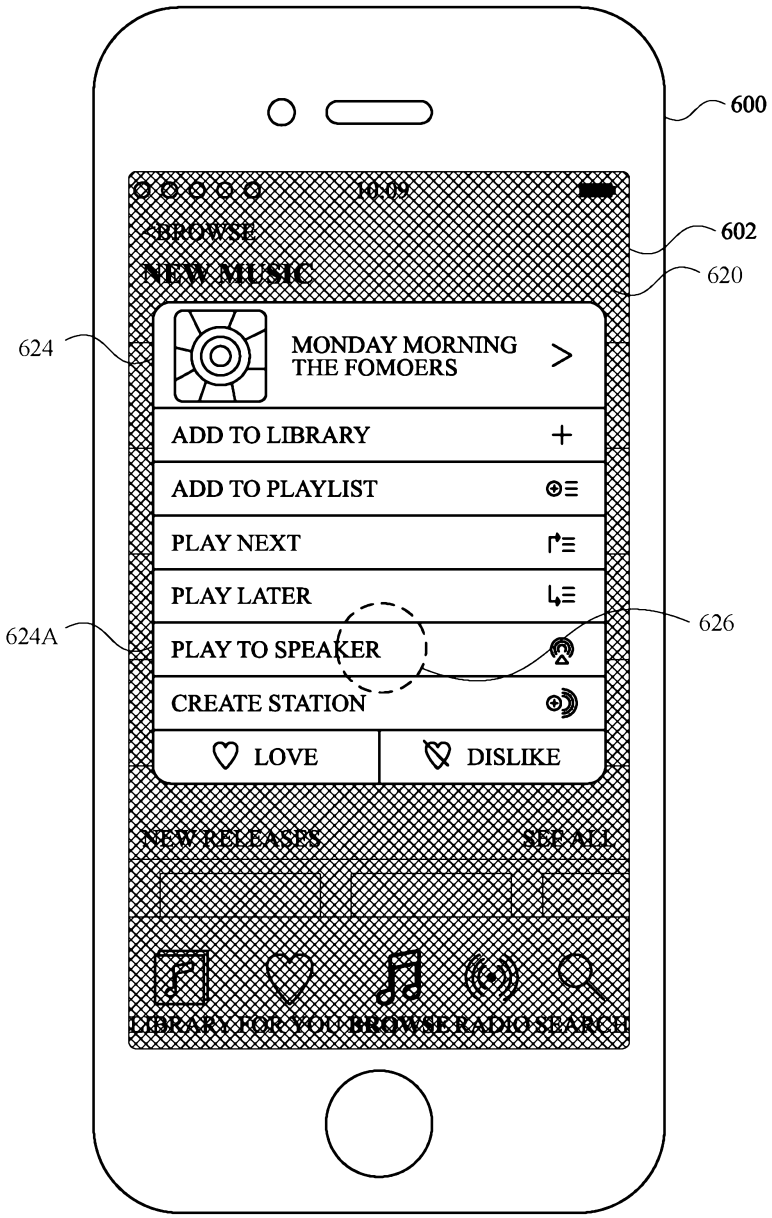

600
10:09
<BROWSE
602
NEW MUSIC
620
624
MONDAY MORNING
THE FOMOERS
>
ADD TO LIBRARY
+
ADD TO PLAYLIST
PLAY NEXT
PLAY LATER
624A
626
PLAY TO SPEAKER
CREATE STATION
LOVE
DISLIKE
NEW RELEASES
SEE ALL
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 6R

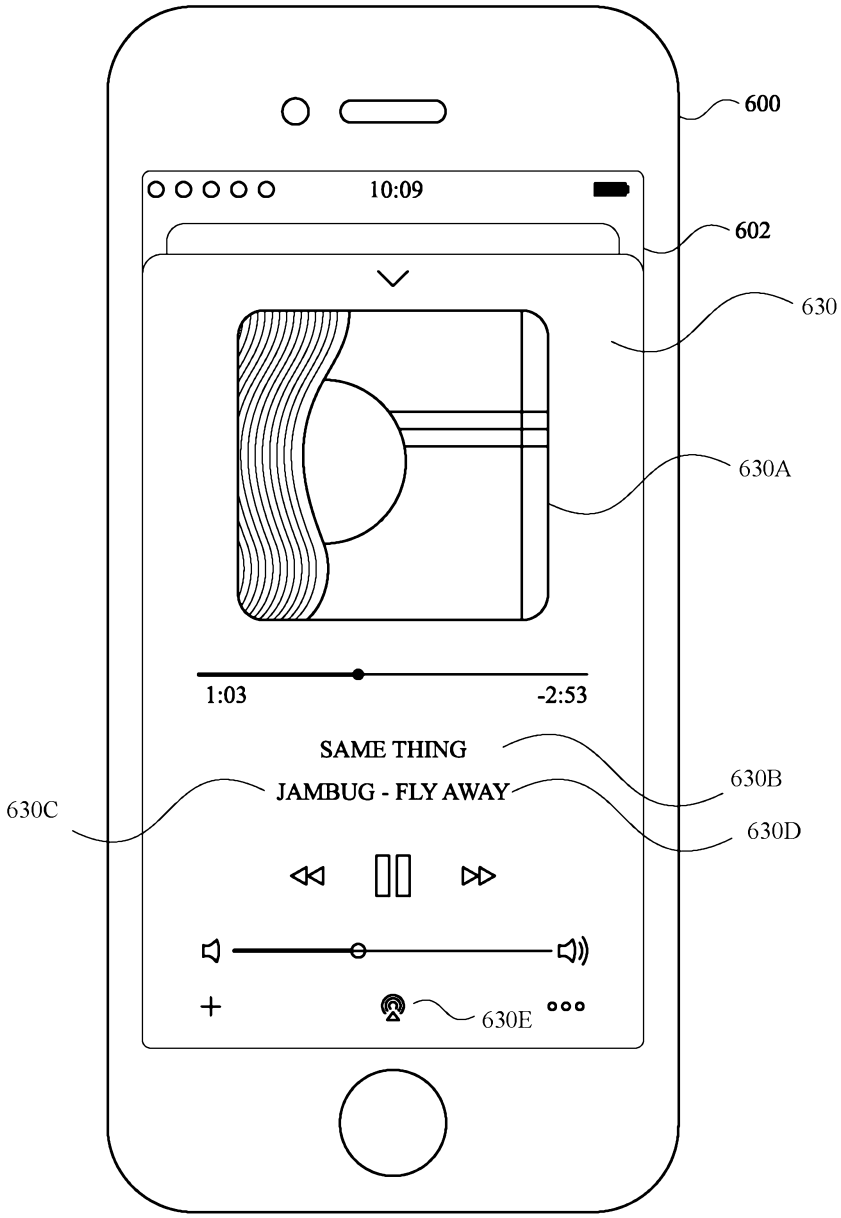

600
10:09
602
630
630A
1:03
-2:53
SAME THING
JAMBUG - FLY AWAY
630B
630C
630D
630E

FIG. 6S

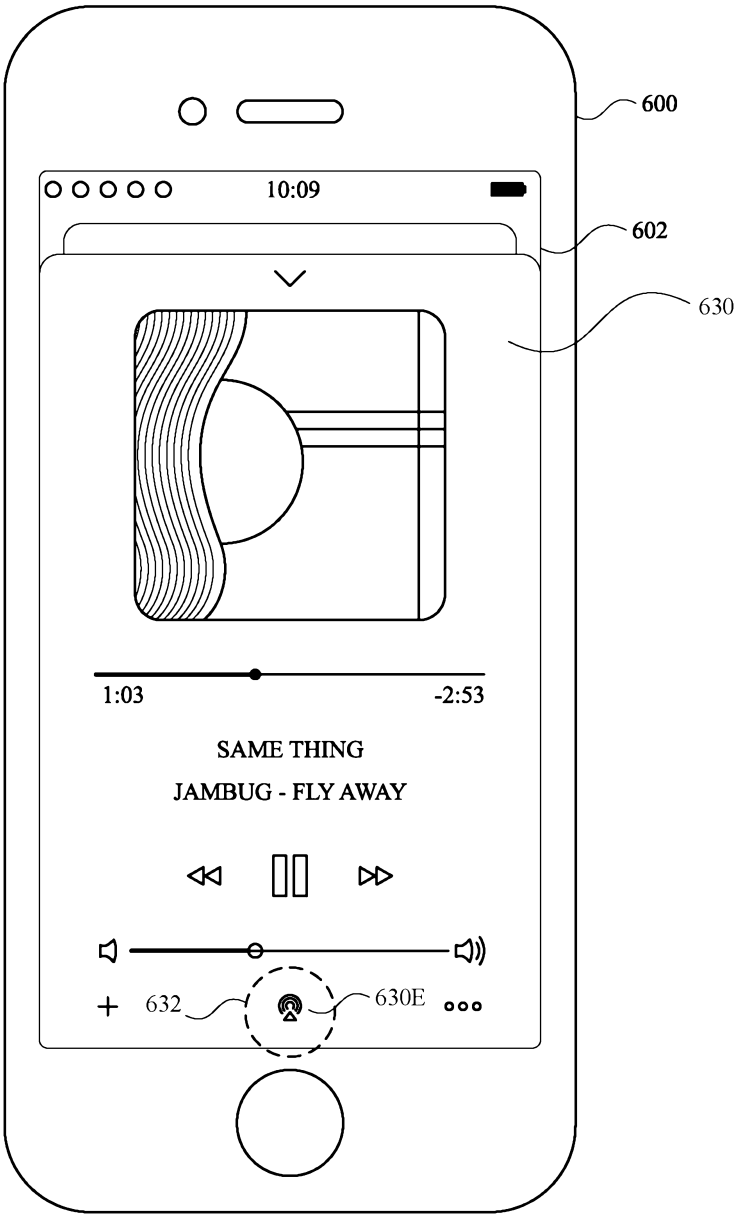


FIG. 6T

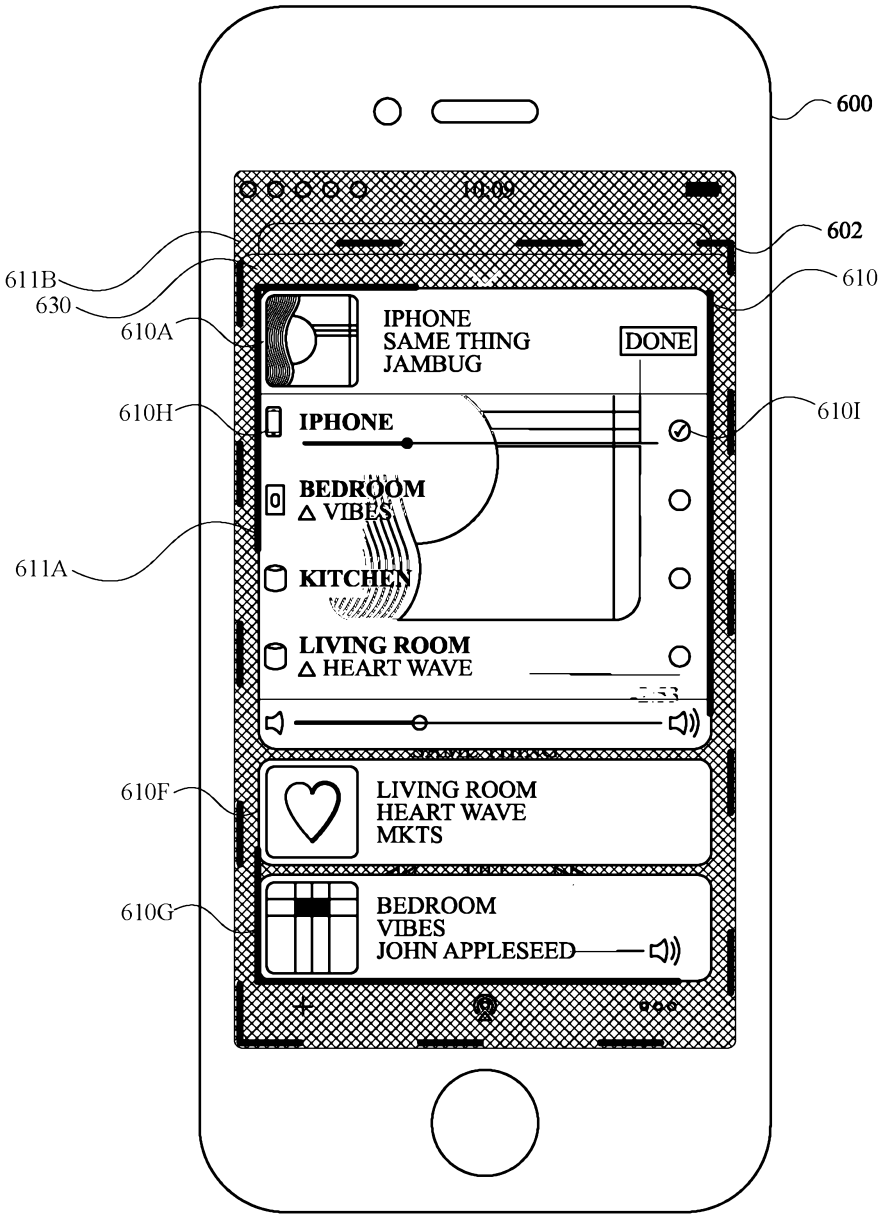

600
602
610
611B
630
610A
610H
610I
611A
610F
610G
10:09
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
BEDROOM
△ VIBES
KITCHEN
LIVING ROOM
△ HEART WAVE
LIVING ROOM
HEART WAVE
MKTS
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6U

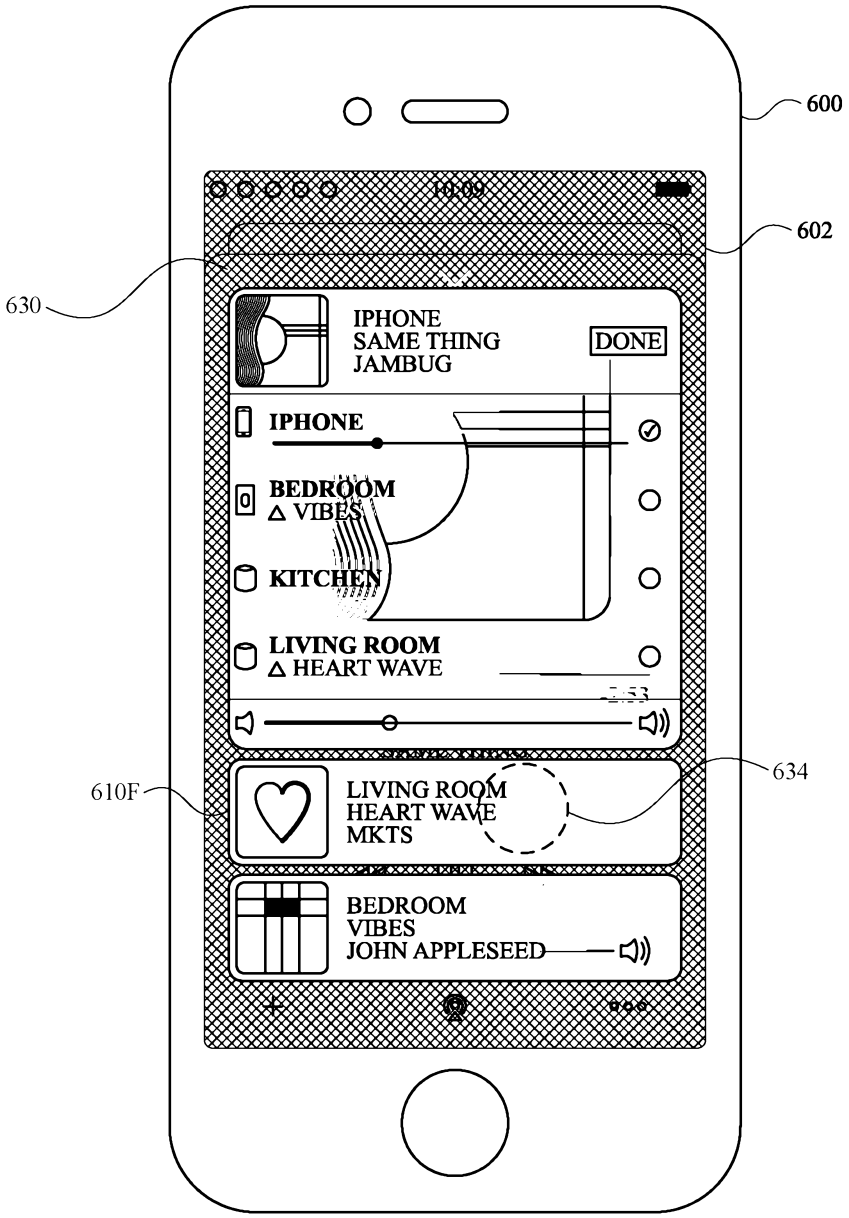

600
602
630
10:09
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
BEDROOM
△ VIBES
KITCHEN
LIVING ROOM
△ HEART WAVE
610F
634
LIVING ROOM
HEART WAVE
MKTS
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6V

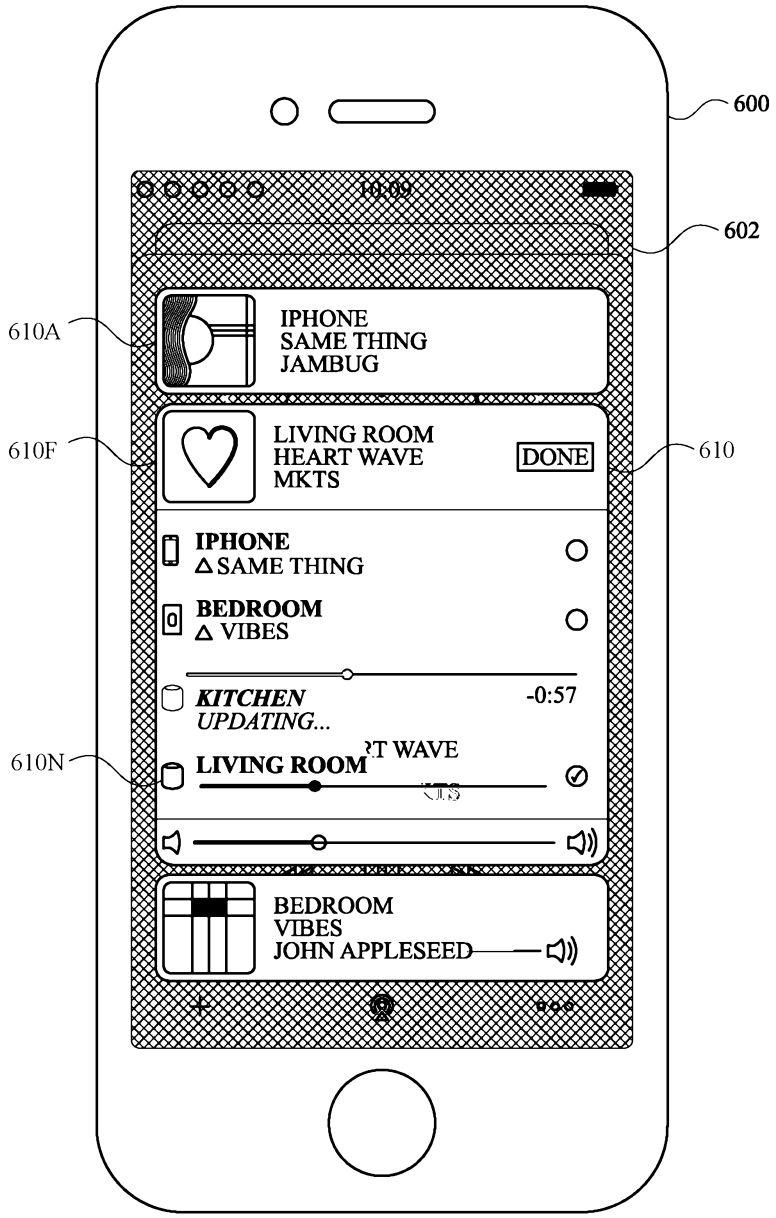

600
602
610A
IPHONE
SAME THING
JAMBUG
610F
LIVING ROOM
HEART WAVE
MKTS
DONE
610
IPHONE
△SAME THING
BEDROOM
△ VIBES
KITCHEN
UPDATING...
-0:57
610N
LIVING ROOM
BEDROOM
VIBES
JOHN APPLESEED

FIG. 6W

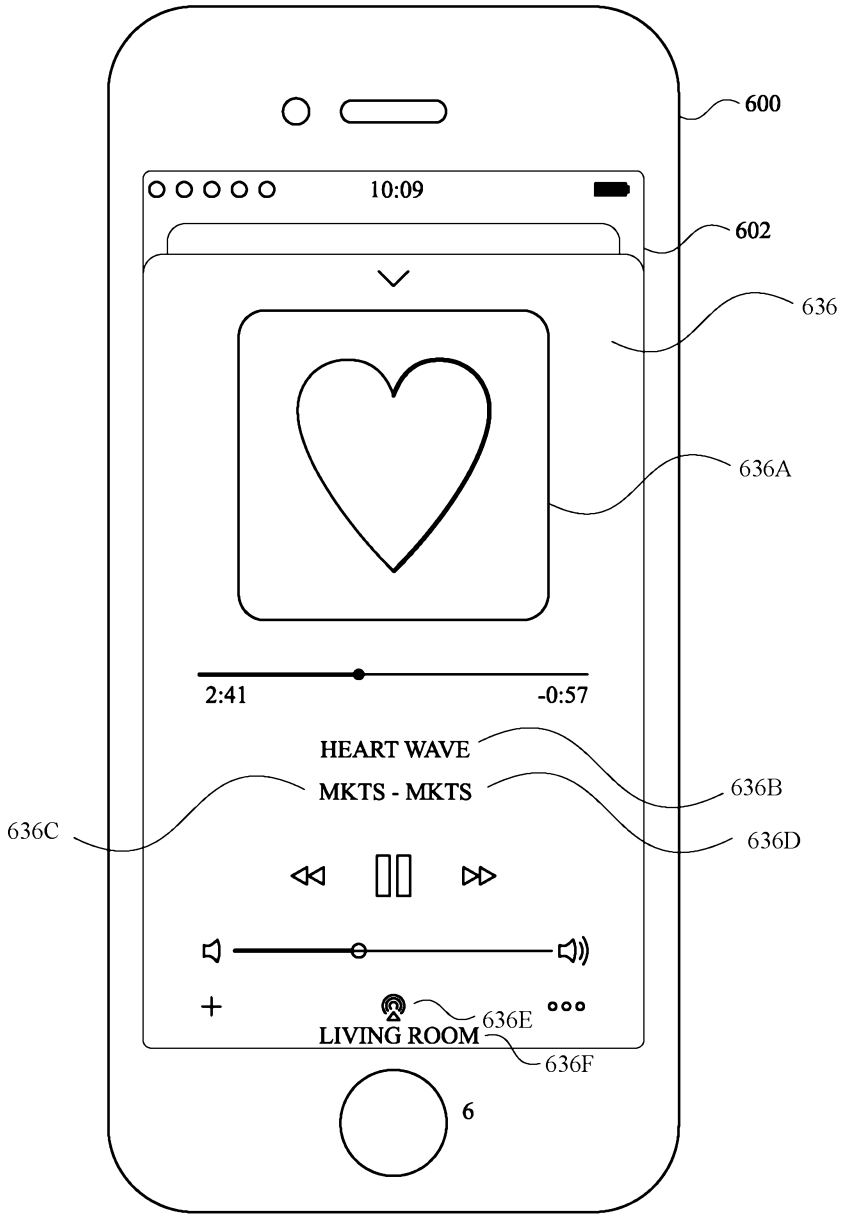

600
10:09
602
636
636A
2:41
-0:57
HEART WAVE
MKTS - MKTS
636B
636C
636D
636E
LIVING ROOM
636F
6

FIG. 6X

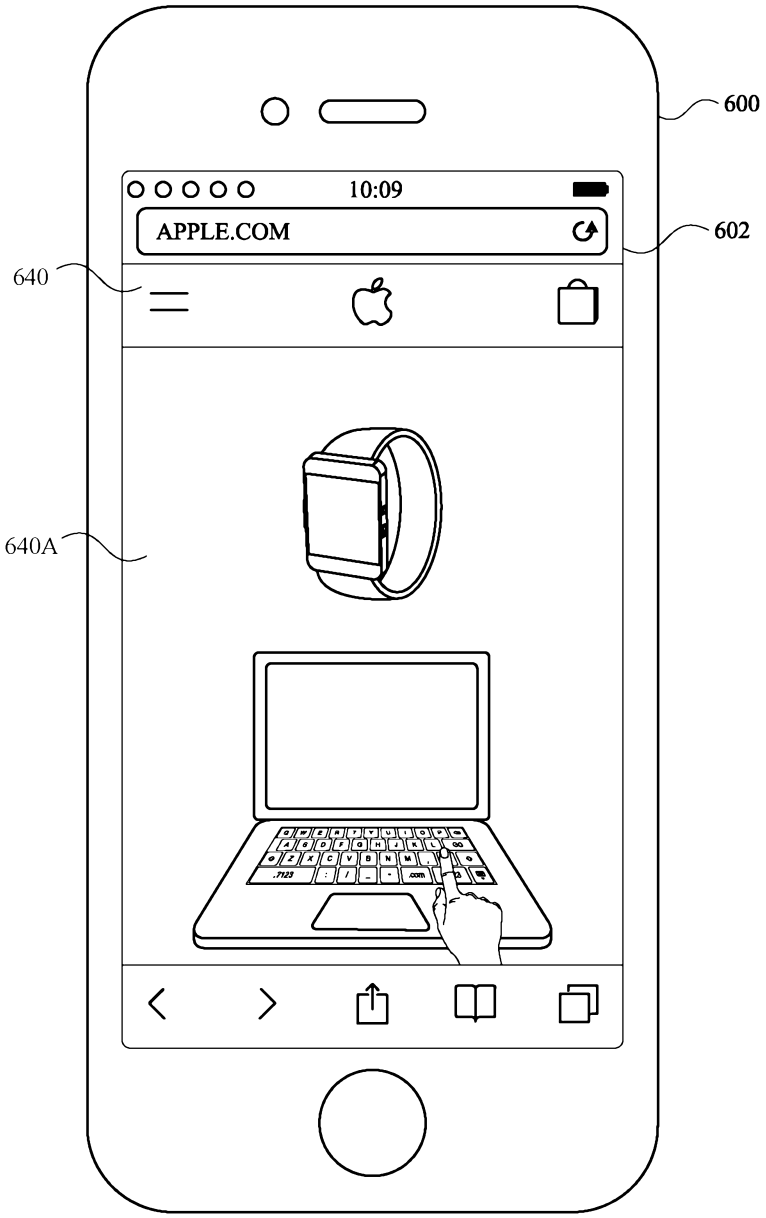


FIG. 6Y

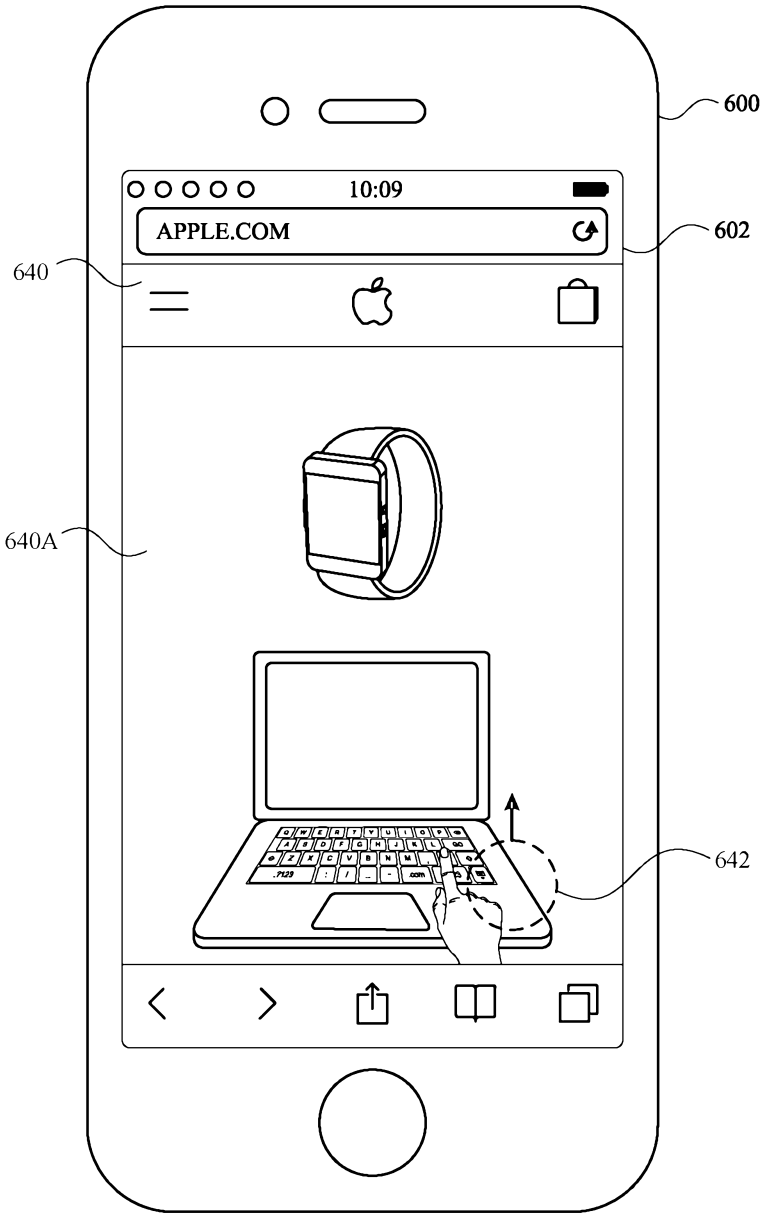


FIG. 6Z

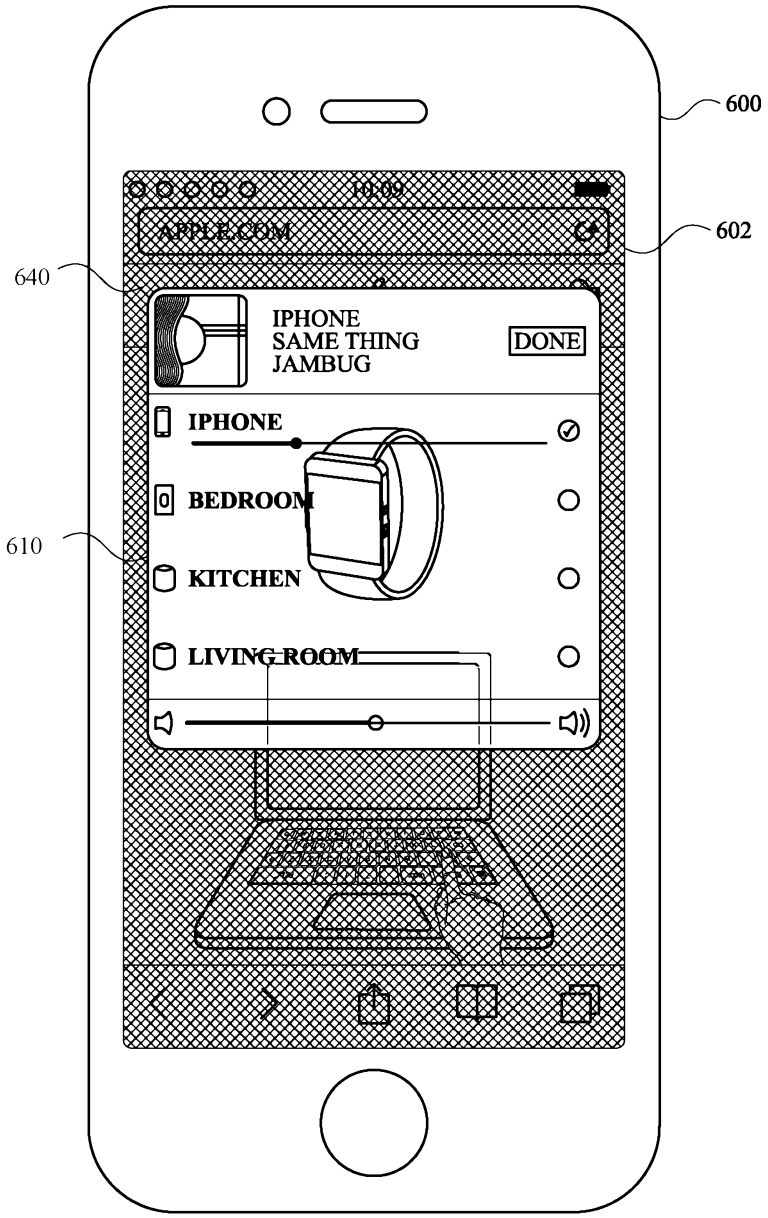

600
602
640
610
10:09
APPLE.COM
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
BEDROOM
KITCHEN
LIVING ROOM

FIG. 6AA

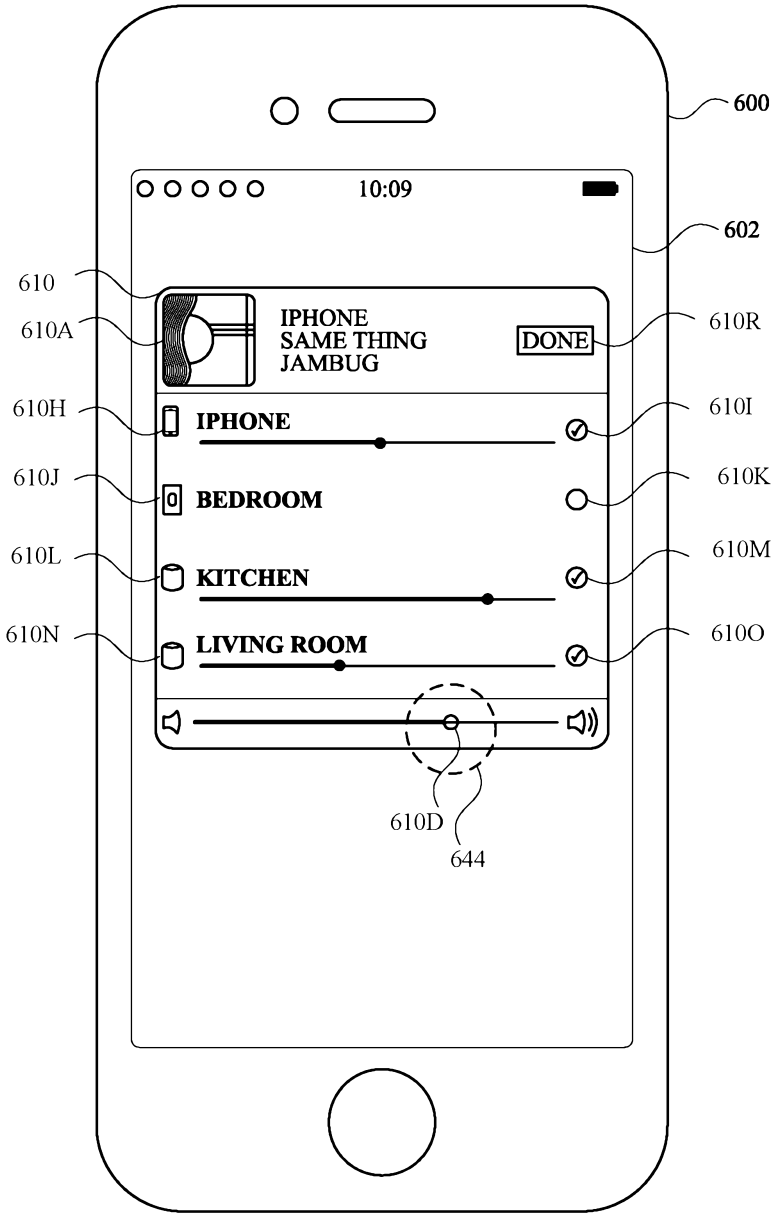

600
10:09
602
610
610A
IPHONE
SAME THING
JAMBUG
DONE
610R
610H
IPHONE
610I
610J
BEDROOM
610K
610L
KITCHEN
610M
610N
LIVING ROOM
610O
610D
644

FIG. 6AB

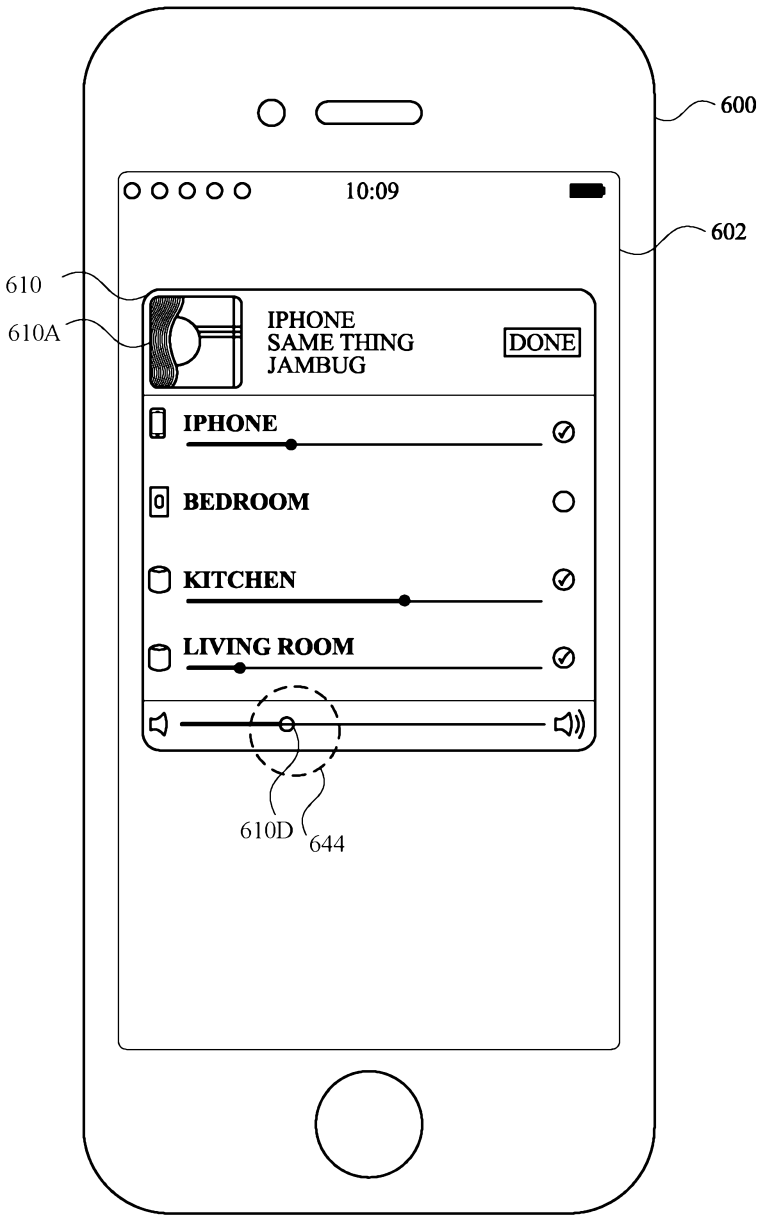

600
10:09
602
610
610A
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
BEDROOM
KITCHEN
LIVING ROOM
610D
644

FIG. 6AC

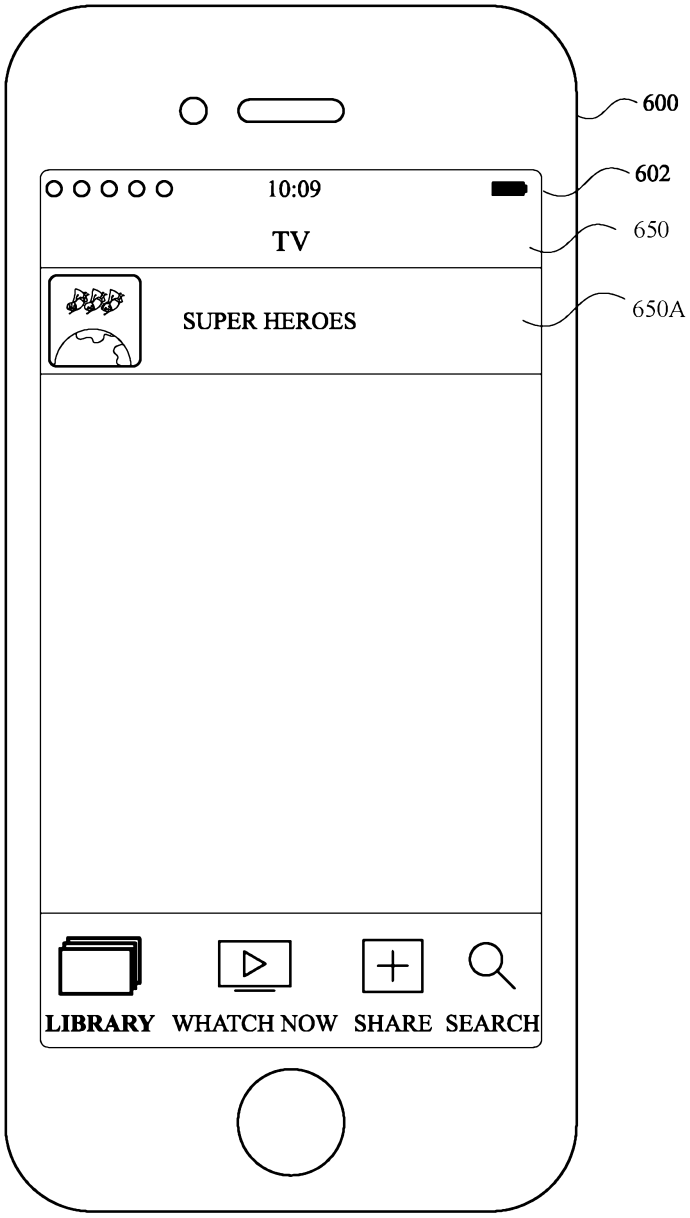

600
602
650
650A
10:09
TV
SUPER HEROES
LIBRARY
WHATCH NOW
SHARE
SEARCH

FIG. 6AD

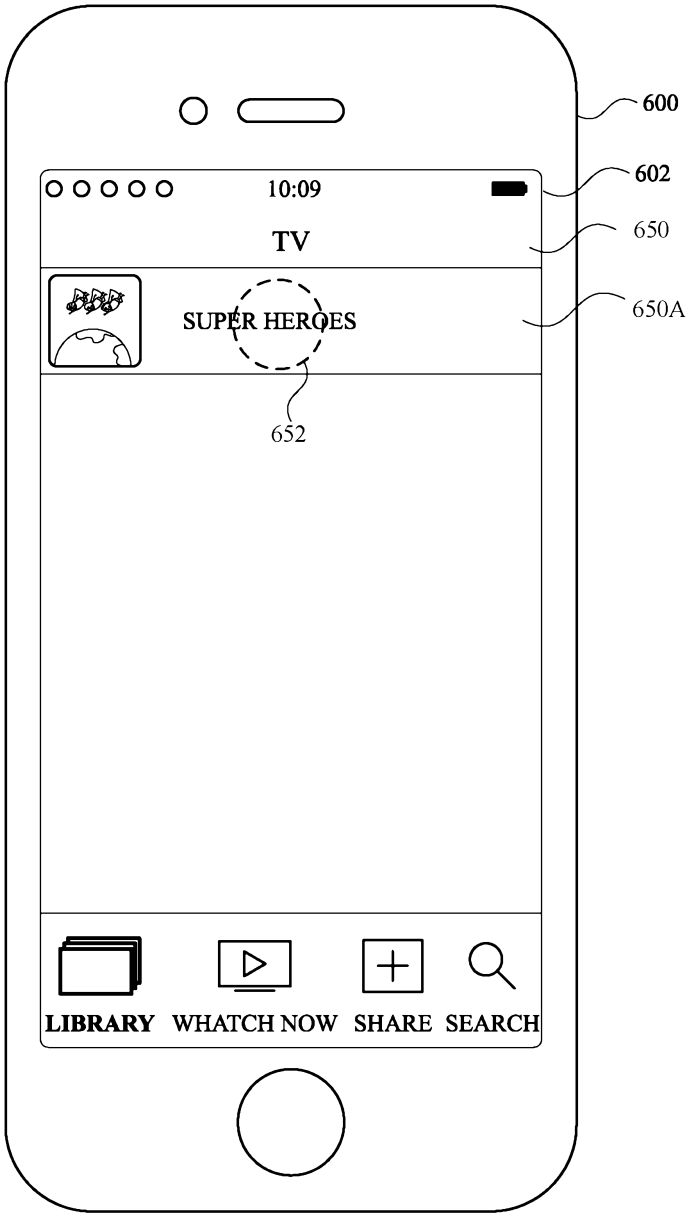

600
602
10:09
TV
650
650A
SUPER HEROES
652
LIBRARY
WHATCH NOW
SHARE
SEARCH

FIG. 6AE

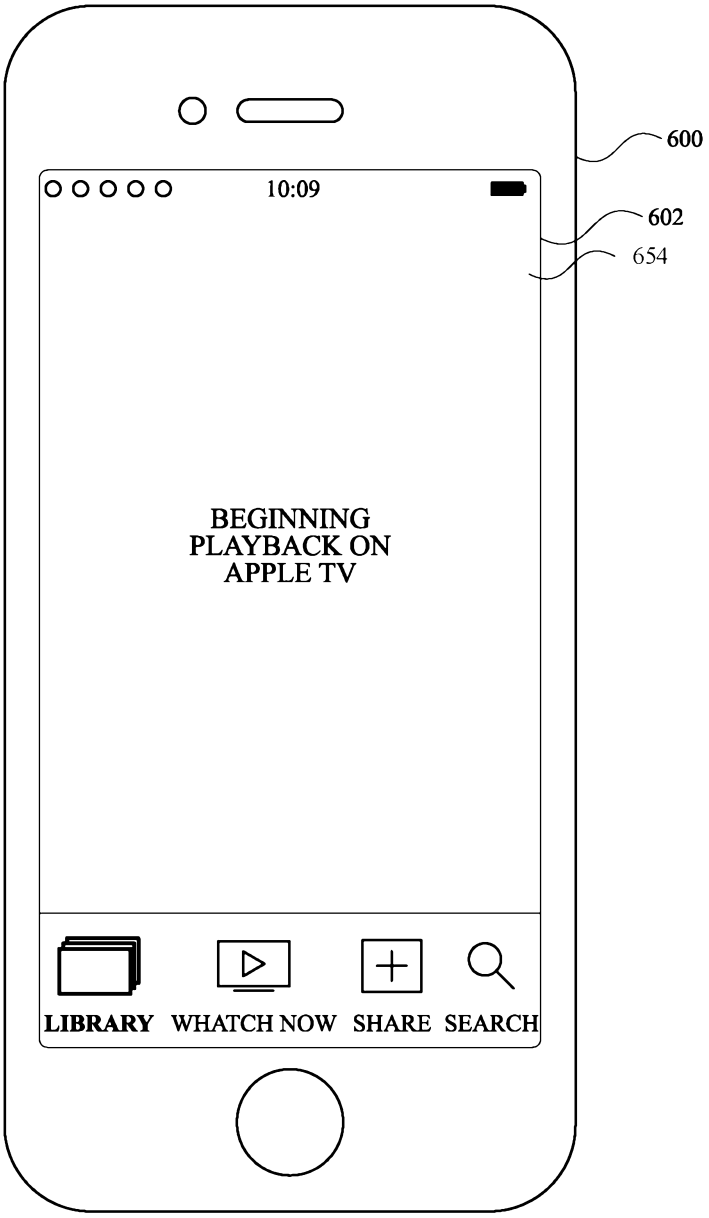

600
10:09
602
654
BEGINNING
PLAYBACK ON
APPLE TV
LIBRARY
WHATCH NOW
SHARE
SEARCH

FIG. 6AF

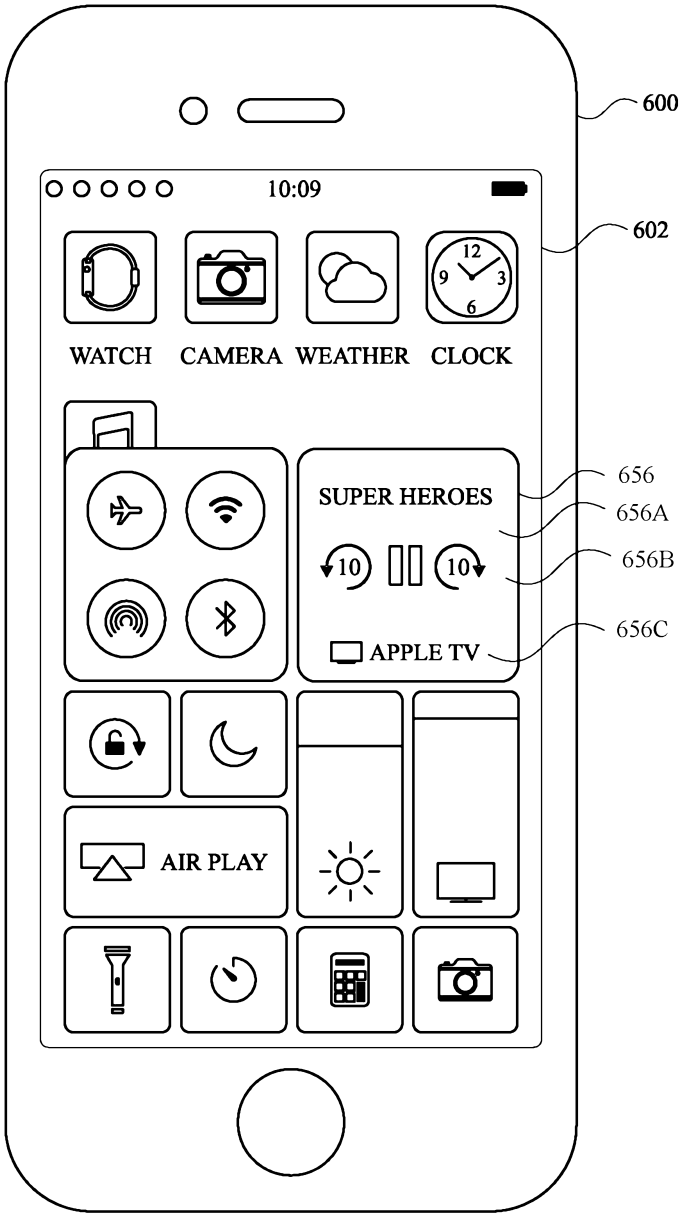

600
10:09
602
WATCH
CAMERA
WEATHER
CLOCK
656
SUPER HEROES
656A
10
10
656B
APPLE TV
656C
AIR PLAY

FIG. 6AG

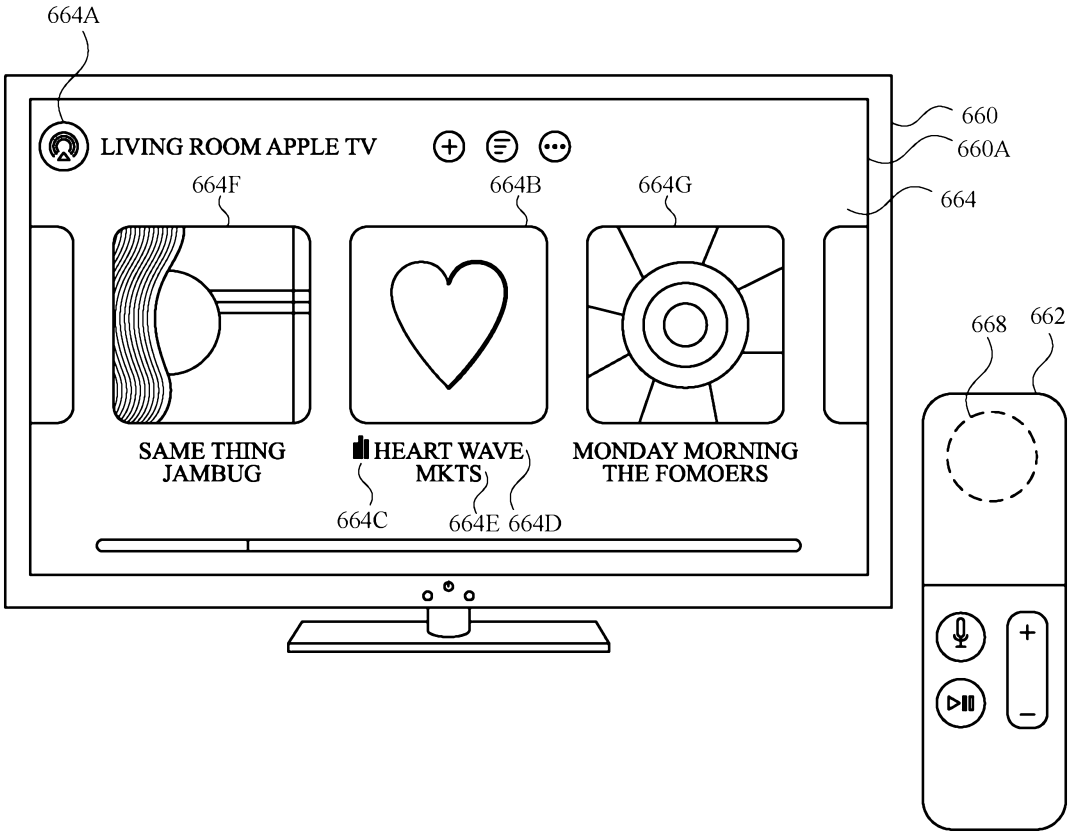

664A
LIVING ROOM APPLE TV
660
660A
664
664F
664B
664G
668
662
SAME THING
JAMBUG
HEART WAVE
MKTS
MONDAY MORNING
THE FOMOERS
664C
664E
664D

FIG. 6AH

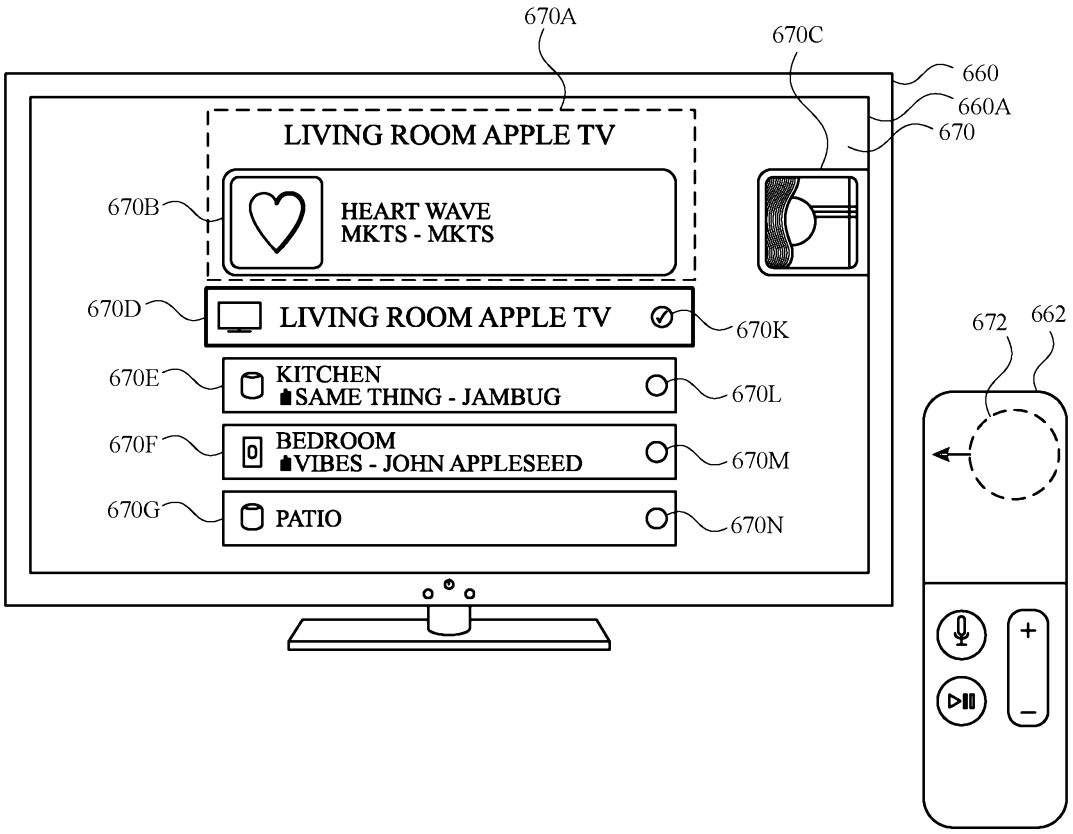

670A
670C
660
660A
670
LIVING ROOM APPLE TV
670B
HEART WAVE
MKTS - MKTS
670D
LIVING ROOM APPLE TV
670K
670E
KITCHEN
SAME THING - JAMBUG
670L
670F
BEDROOM
VIBES - JOHN APPLESEED
670M
670G
PATIO
670N
672
662

FIG. 6AI

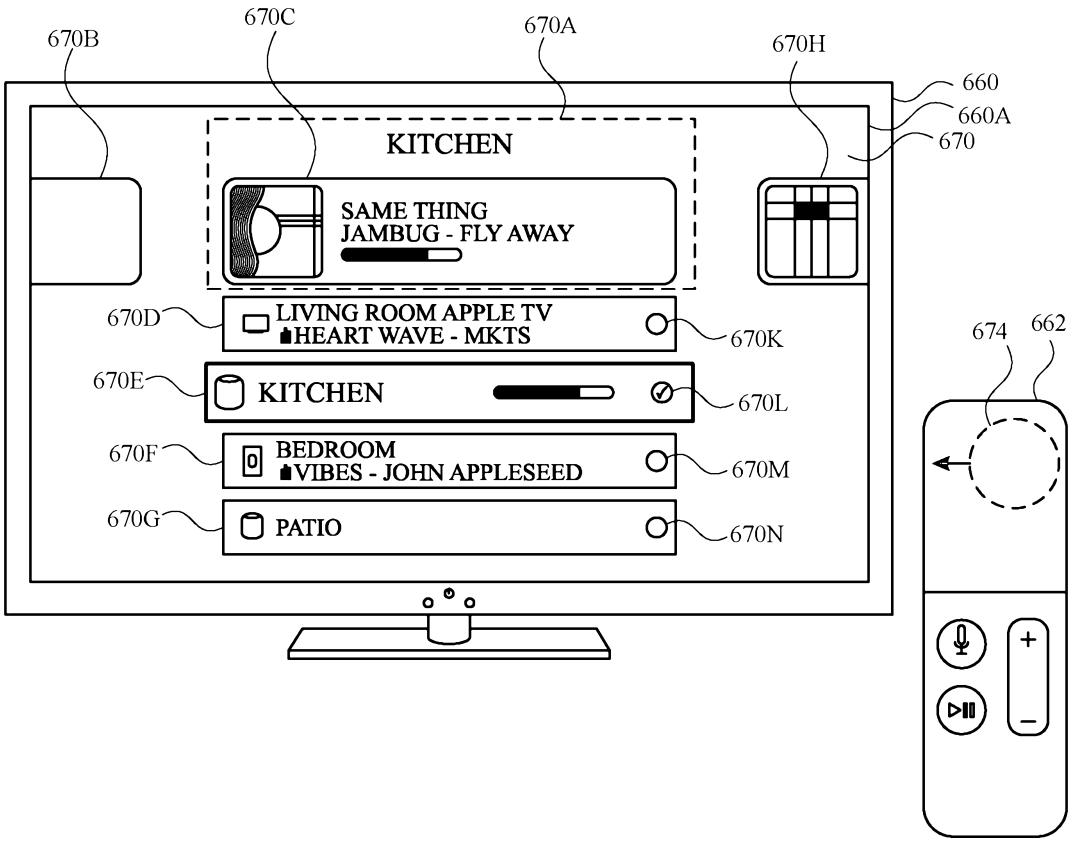

670B
670C
670A
670H
660
660A
670
KITCHEN
SAME THING
JAMBUG - FLY AWAY
670D
LIVING ROOM APPLE TV
HEART WAVE - MKTS
670K
670E
KITCHEN
670L
670F
BEDROOM
VIBES - JOHN APPLESEED
670M
670G
PATIO
670N
674
662

FIG. 6AJ

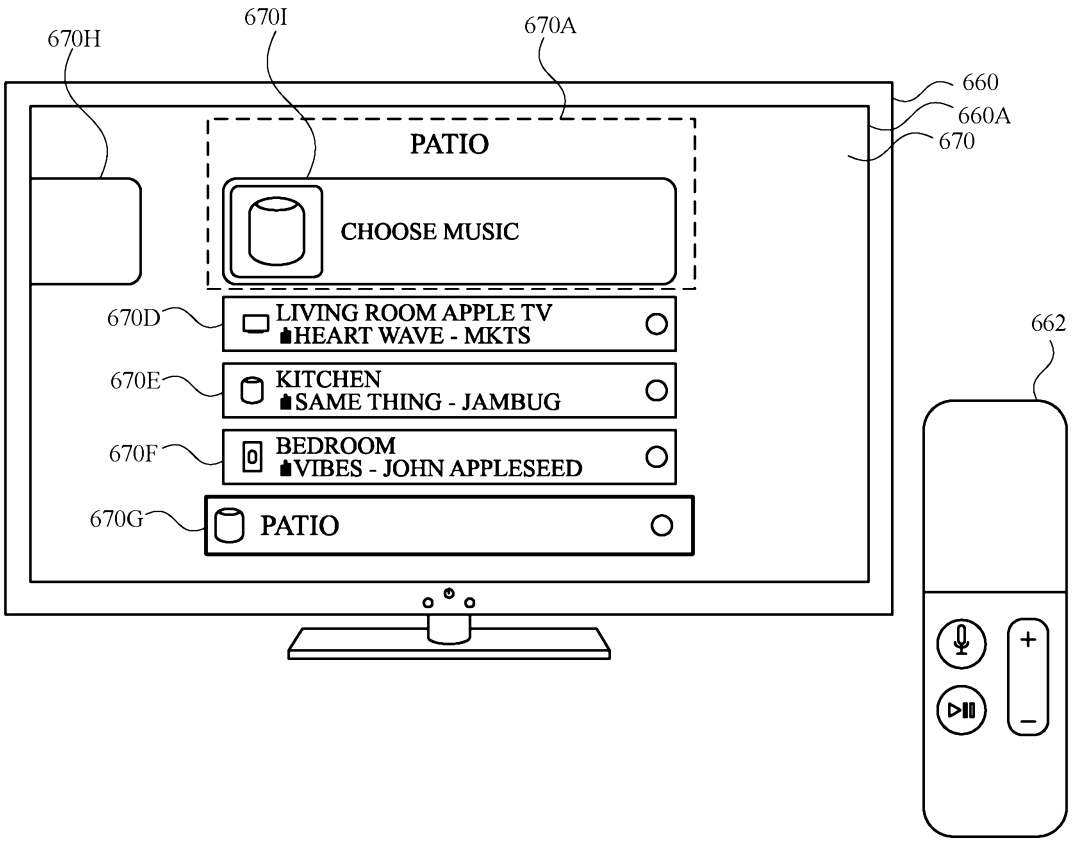

670H
670I
670A
660
660A
670
PATIO
CHOOSE MUSIC
670D
LIVING ROOM APPLE TV
HEART WAVE - MKTS
670E
KITCHEN
SAME THING - JAMBUG
670F
BEDROOM
VIBES - JOHN APPLESEED
670G
PATIO
662

FIG. 6AK

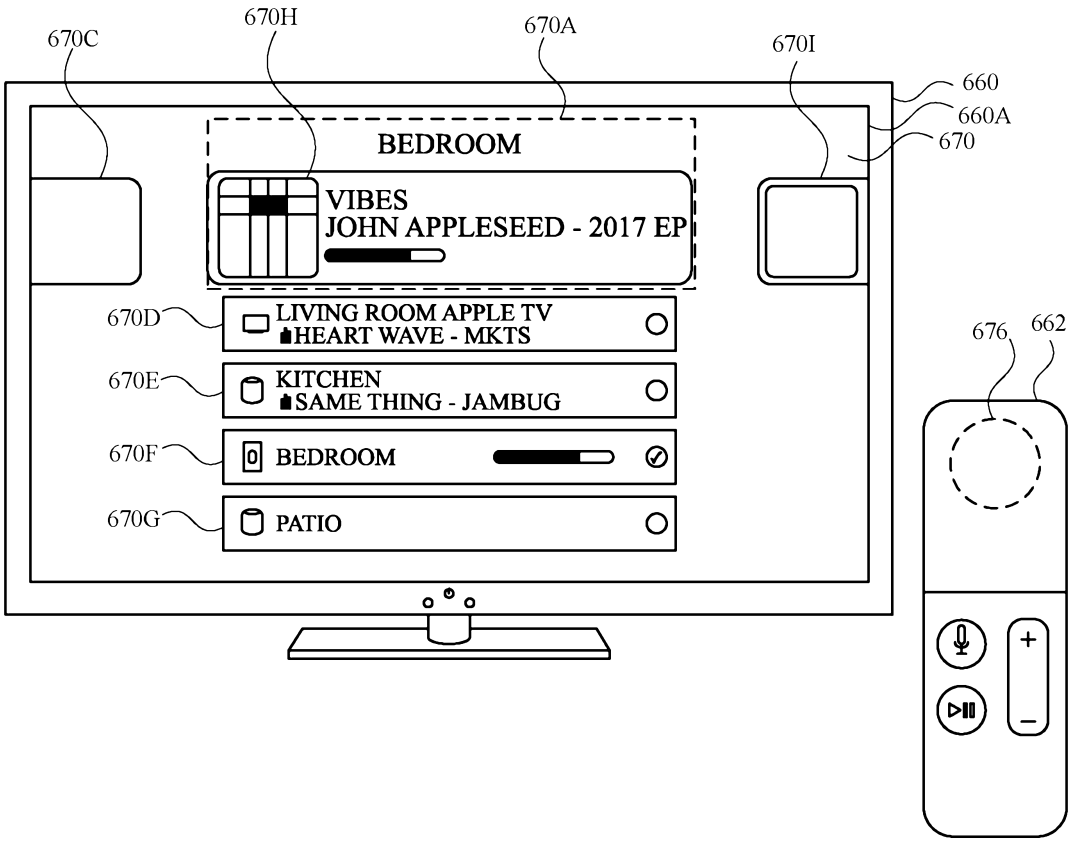

670C
670H
670A
670I
660
660A
670
BEDROOM
VIBES
JOHN APPLESEED - 2017 EP
670D
LIVING ROOM APPLE TV
HEART WAVE - MKTS
670E
KITCHEN
SAME THING - JAMBUG
670F
BEDROOM
670G
PATIO
676
662

FIG. 6AL

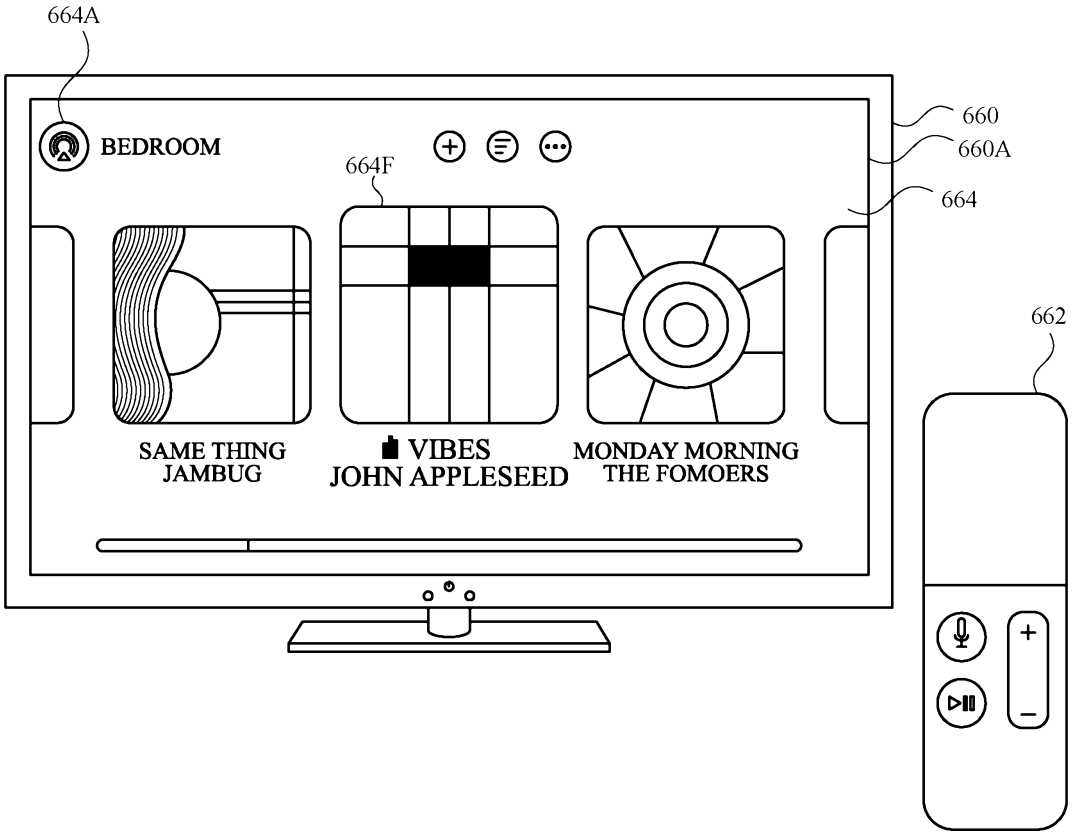

664A
BEDROOM
664F
660
660A
664
662
SAME THING
JAMBUG
VIBES
JOHN APPLESEED
MONDAY MORNING
THE FOMOERS

FIG. 6AM

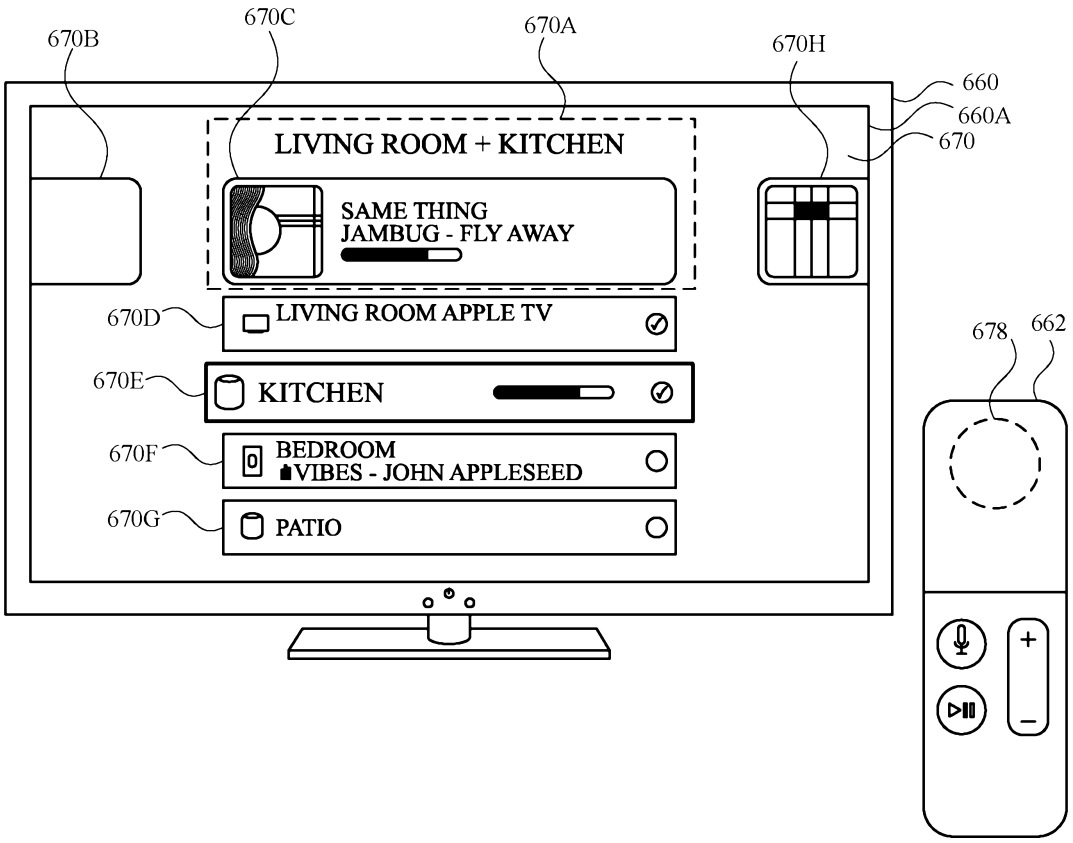

670B
670C
670A
670H
660
660A
670
LIVING ROOM + KITCHEN
SAME THING
JAMBUG - FLY AWAY
670D
LIVING ROOM APPLE TV
670E
KITCHEN
670F
BEDROOM
VIBES - JOHN APPLESEED
670G
PATIO
678
662

FIG. 6AN

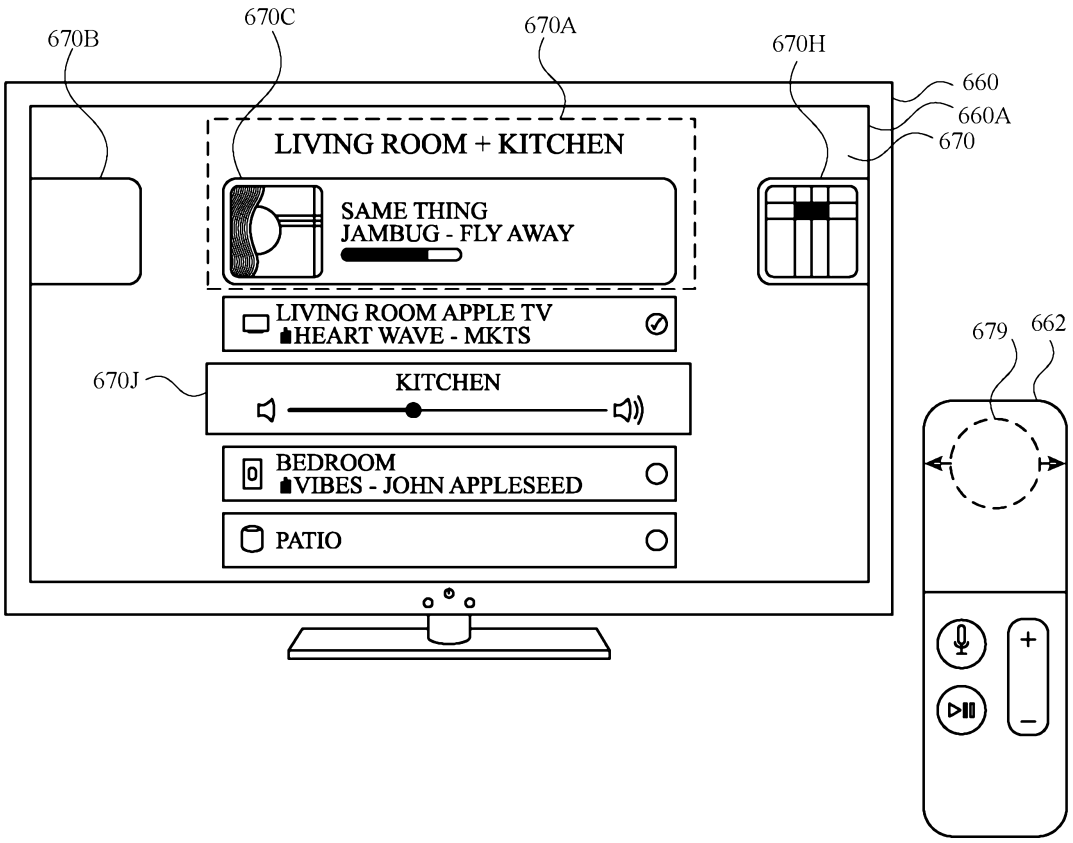

670B
670C
670A
670H
660
660A
670
LIVING ROOM + KITCHEN
SAME THING
JAMBUG - FLY AWAY
LIVING ROOM APPLE TV
HEART WAVE - MKTS
670J
KITCHEN
BEDROOM
VIBES - JOHN APPLESEED
PATIO
679
662

FIG. 6AO

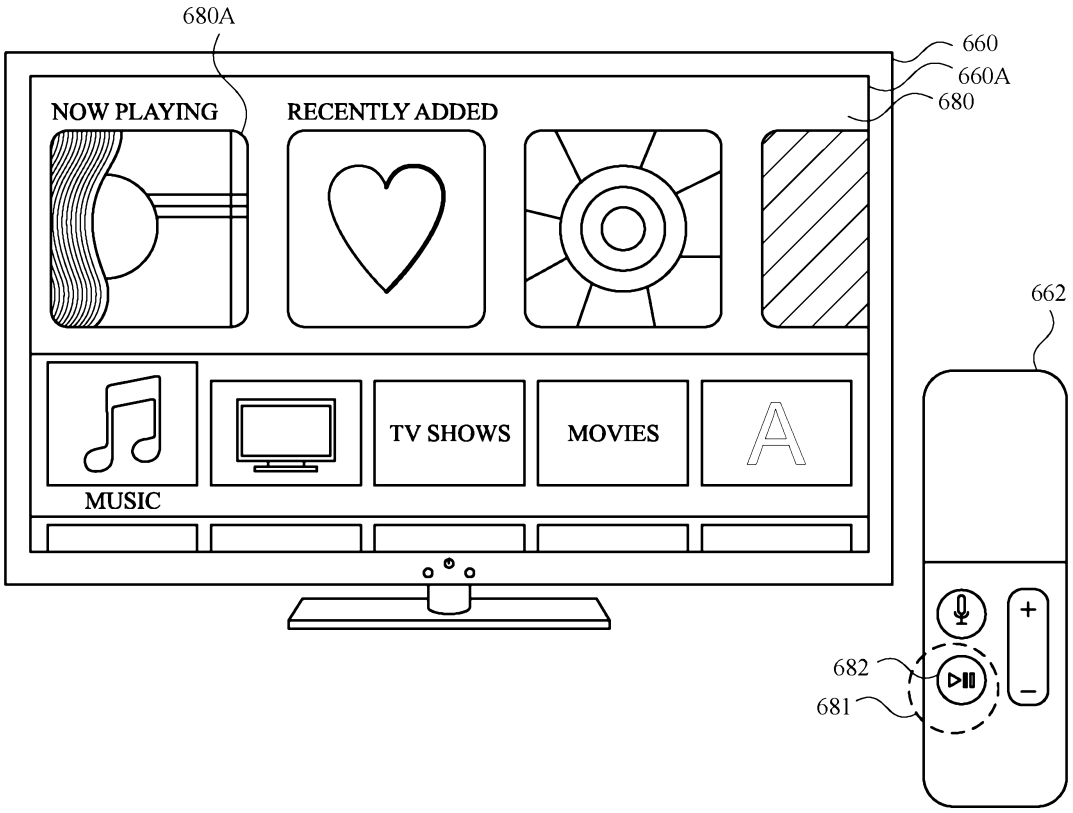

680A
660
660A
680
NOW PLAYING
RECENTLY ADDED
662
TV SHOWS
MOVIES
MUSIC
682
681

FIG. 6AP

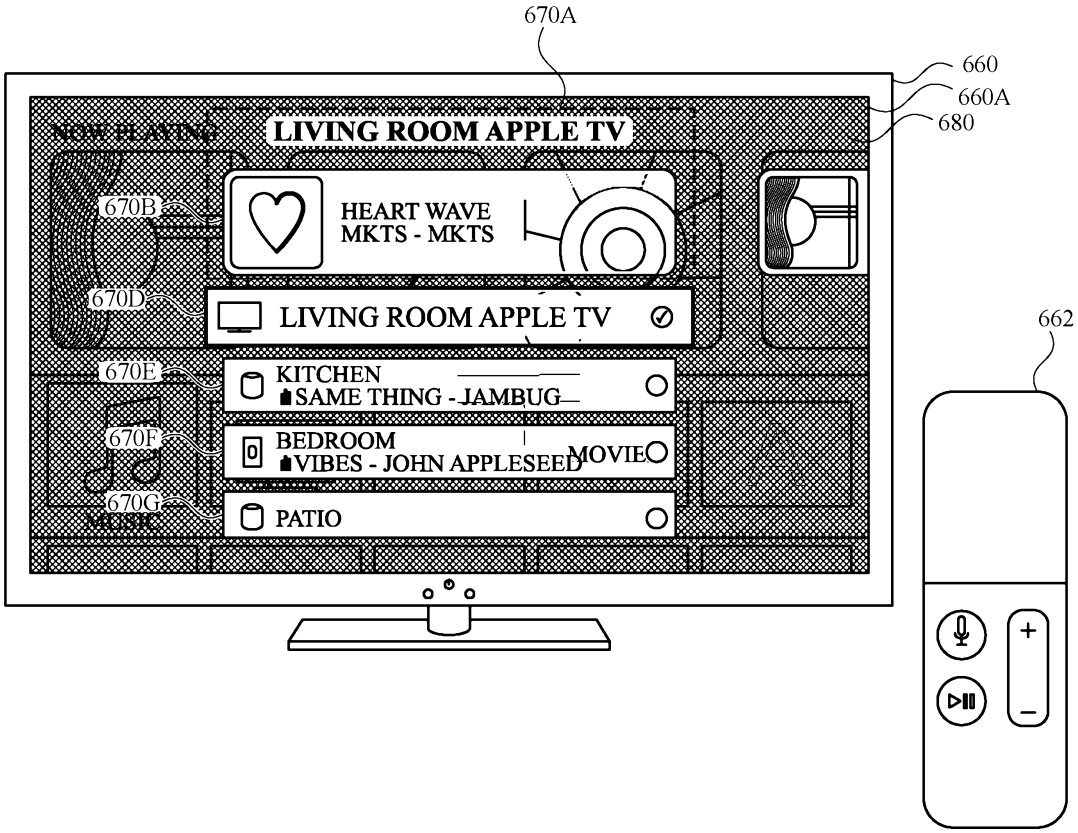

670A
660
660A
680
NOW PLAYING
LIVING ROOM APPLE TV
670B
HEART WAVE
MKTS - MKTS
670D
LIVING ROOM APPLE TV
662
670E
KITCHEN
SAME THING - JAMBUG
670F
BEDROOM
VIBES - JOHN APPLESEED
MOVIE
670G
MUSIC
PATIO

FIG. 6AQ

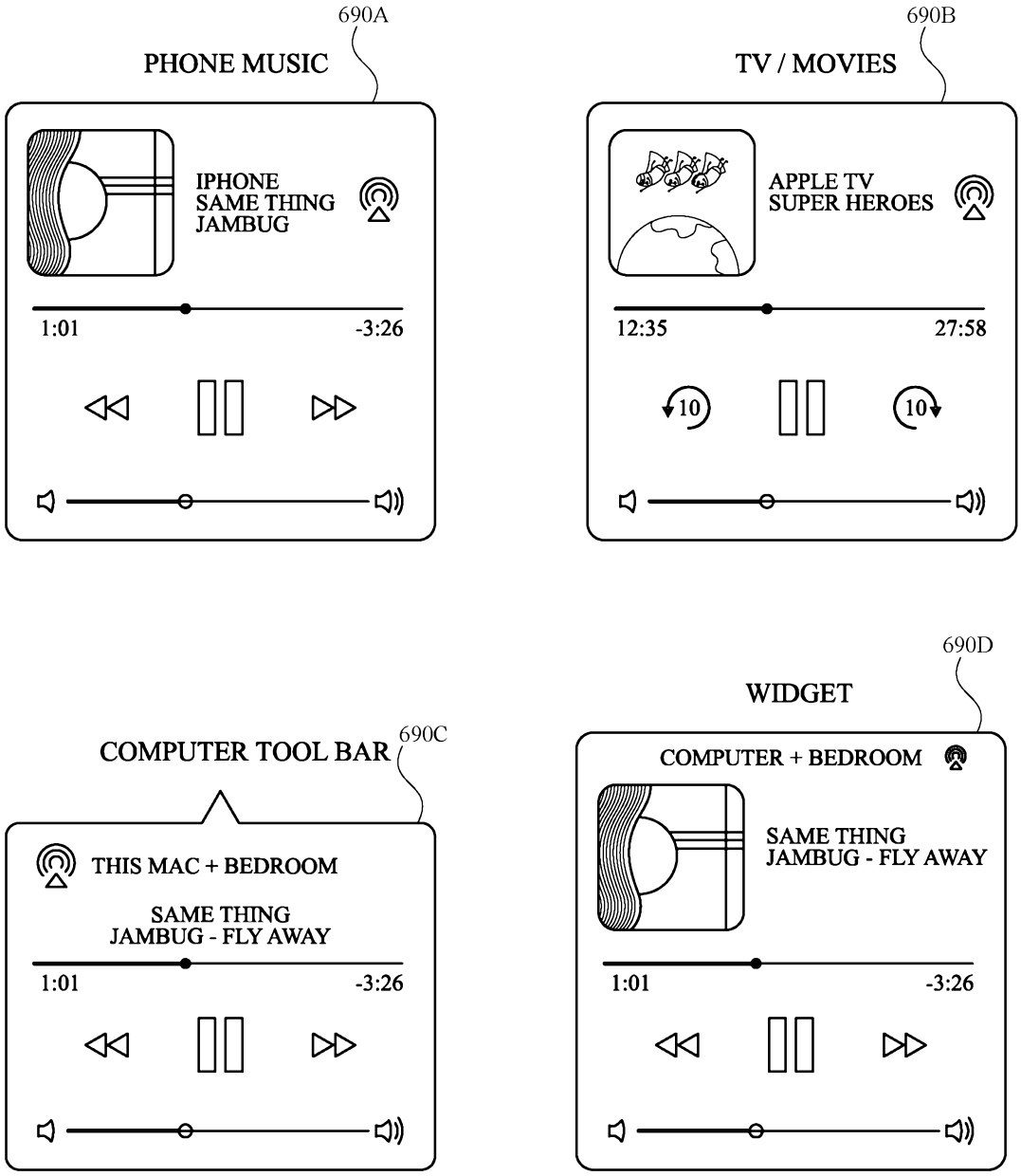

690A
PHONE MUSIC
IPHONE
SAME THING
JAMBUG
1:01
-3:26
690B
TV / MOVIES
APPLE TV
SUPER HEROES
12:35
27:58
10
10
690C
COMPUTER TOOL BAR
THIS MAC + BEDROOM
SAME THING
JAMBUG - FLY AWAY
1:01
-3:26
690D
WIDGET
COMPUTER + BEDROOM
SAME THING
JAMBUG - FLY AWAY
1:01
-3:26

FIG. 6AR

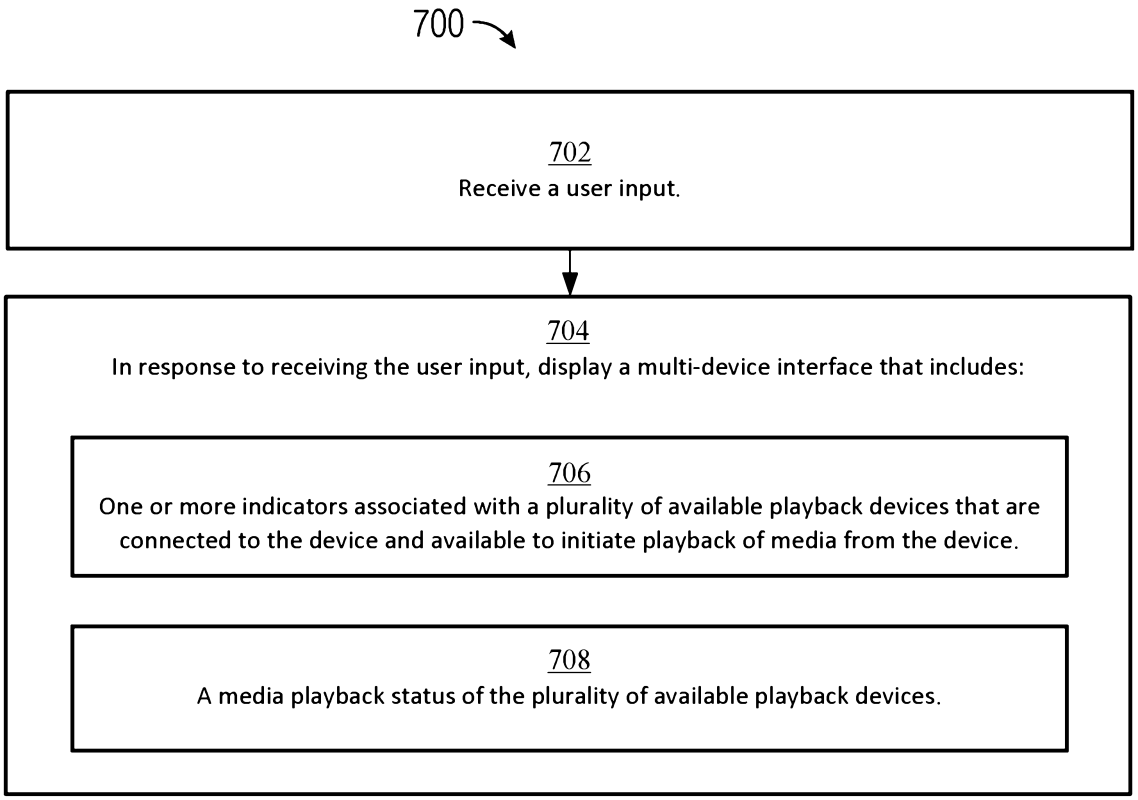

700
702
Receive a user input.
704
In response to receiving the user input, display a multi-device interface that includes:
706
One or more indicators associated with a plurality of available playback devices that are connected to the device and available to initiate playback of media from the device.
708
A media playback status of the plurality of available playback devices.

FIG. 7

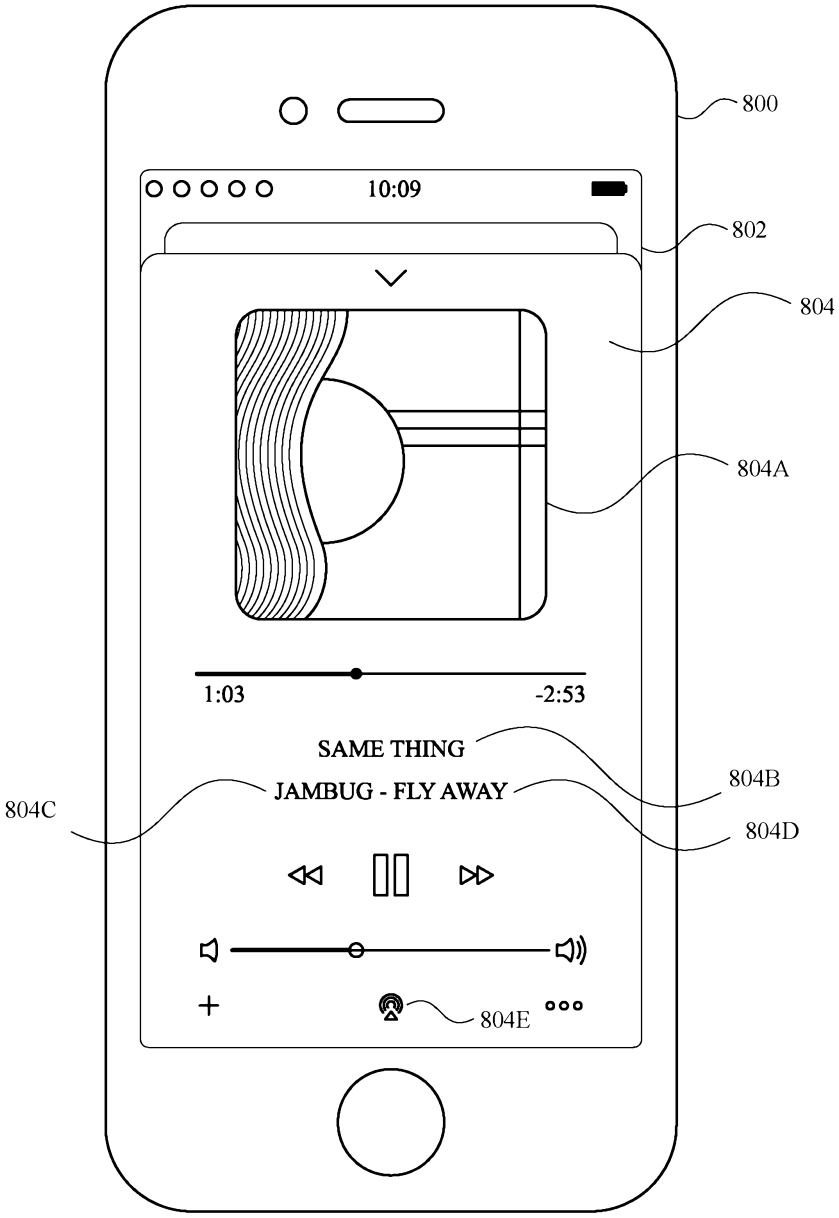

800
10:09
802
804
804A
1:03
-2:53
SAME THING
804B
804C
JAMBUG - FLY AWAY
804D
804E

FIG. 8A

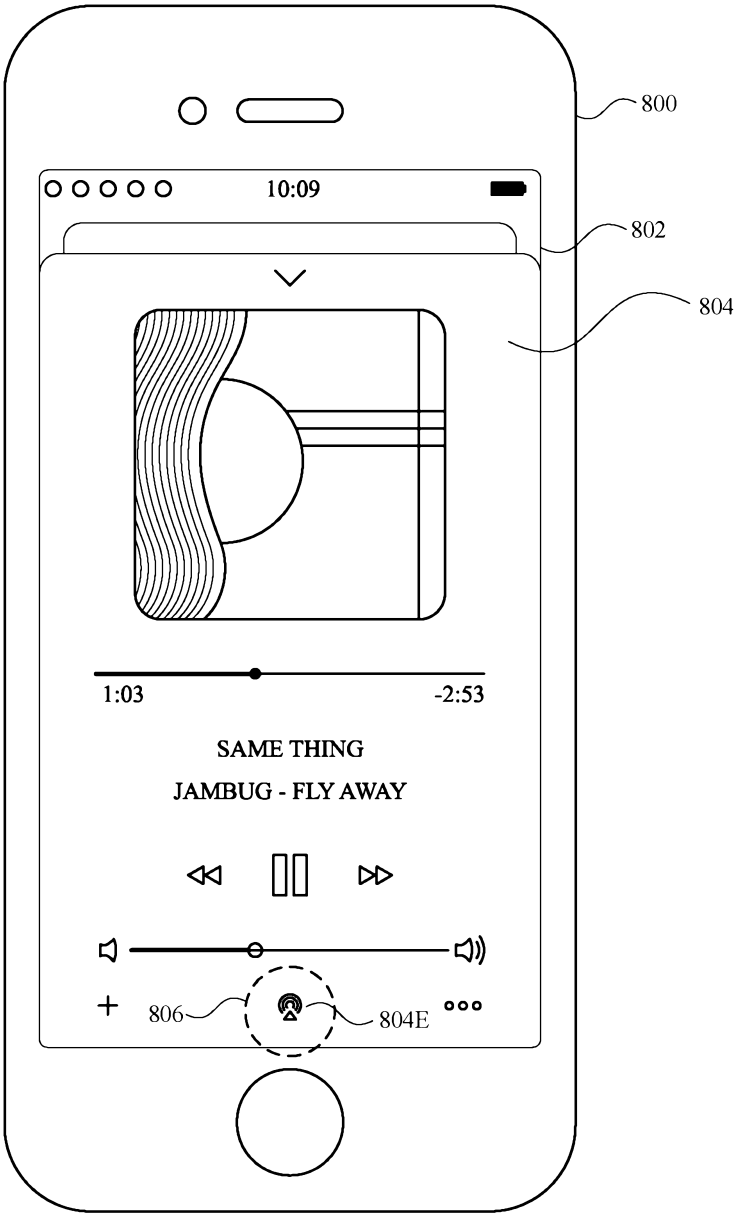


FIG. 8B

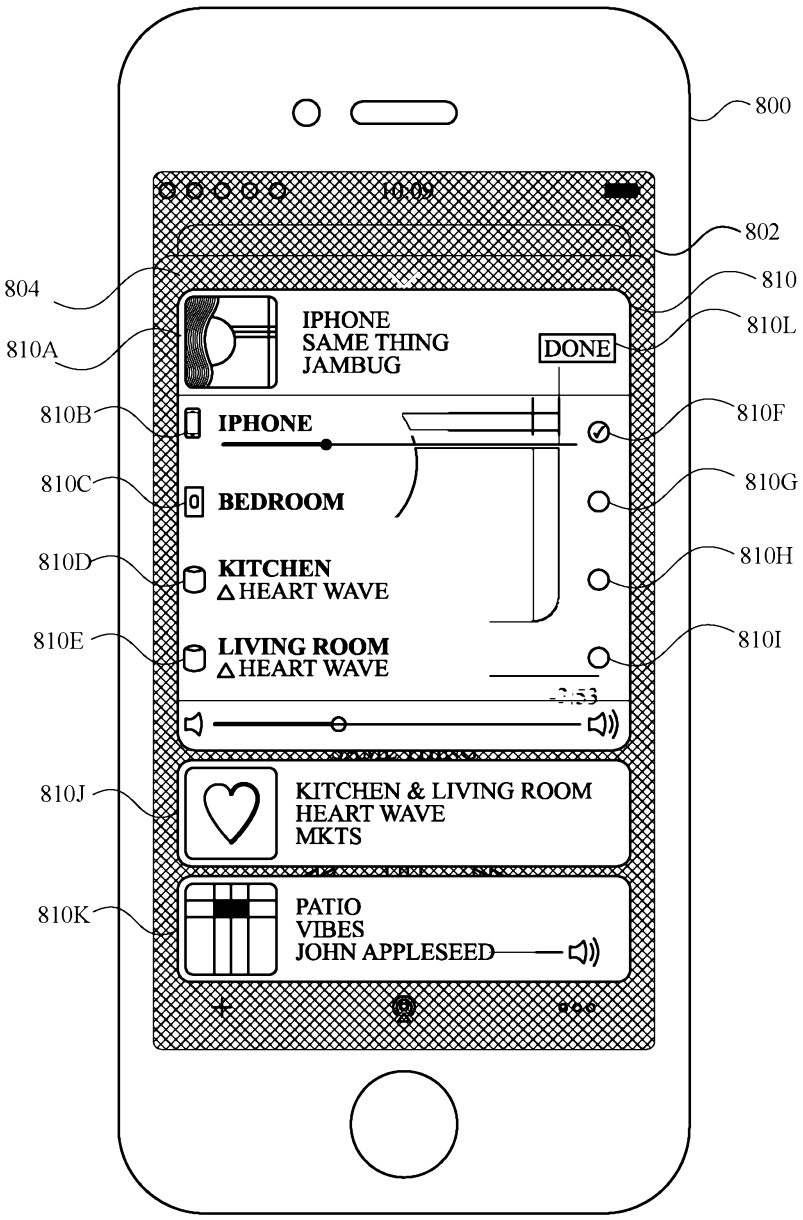

800
10:09
802
804
810
IPHONE
SAME THING
JAMBUG
DONE
810L
810A
810B
IPHONE
810F
810C
BEDROOM
810G
810D
KITCHEN
△HEART WAVE
810H
810E
LIVING ROOM
△HEART WAVE
810I
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
810J
PATIO
VIBES
JOHN APPLESEED
810K

FIG. 8C

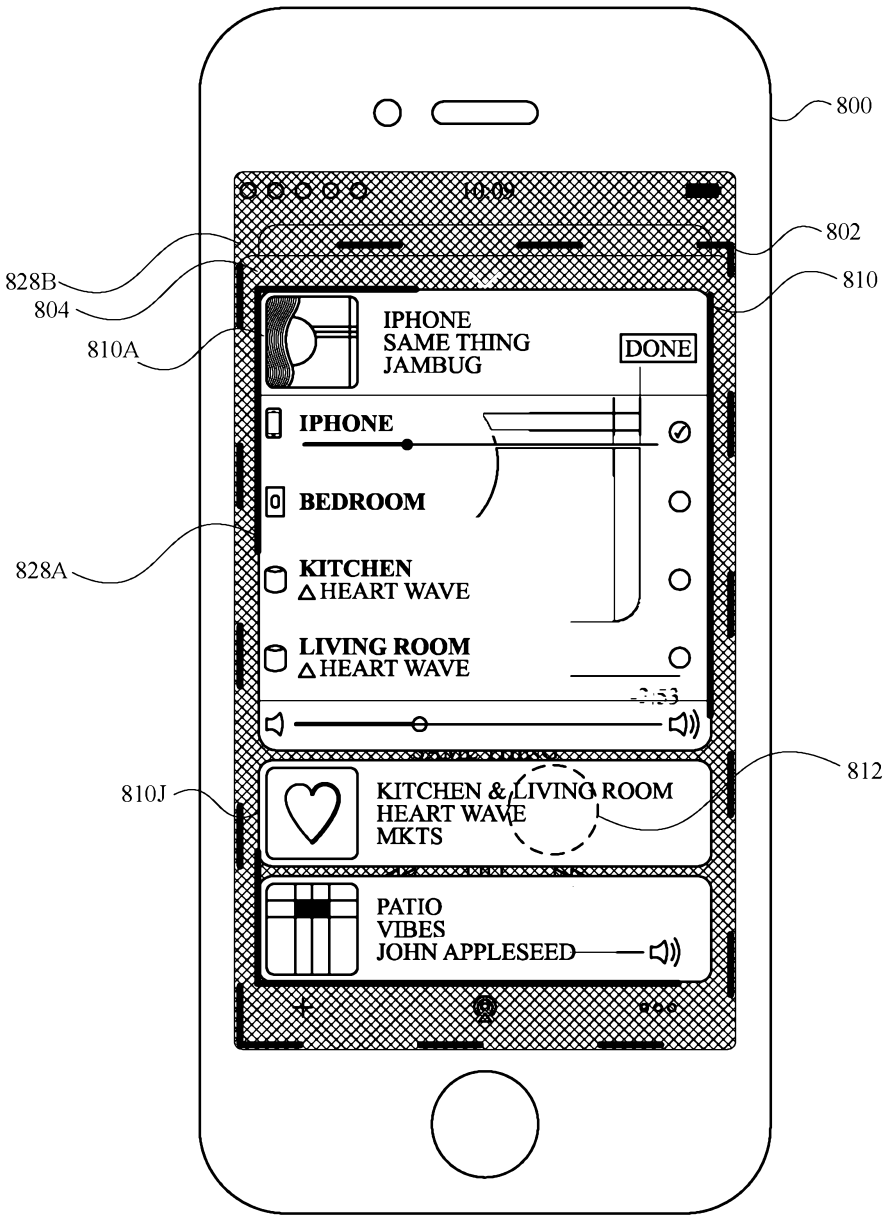

800
802
828B
804
810
810A
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
BEDROOM
828A
KITCHEN
△HEART WAVE
LIVING ROOM
△HEART WAVE
812
810J
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
PATIO
VIBES
JOHN APPLESEED

FIG. 8D

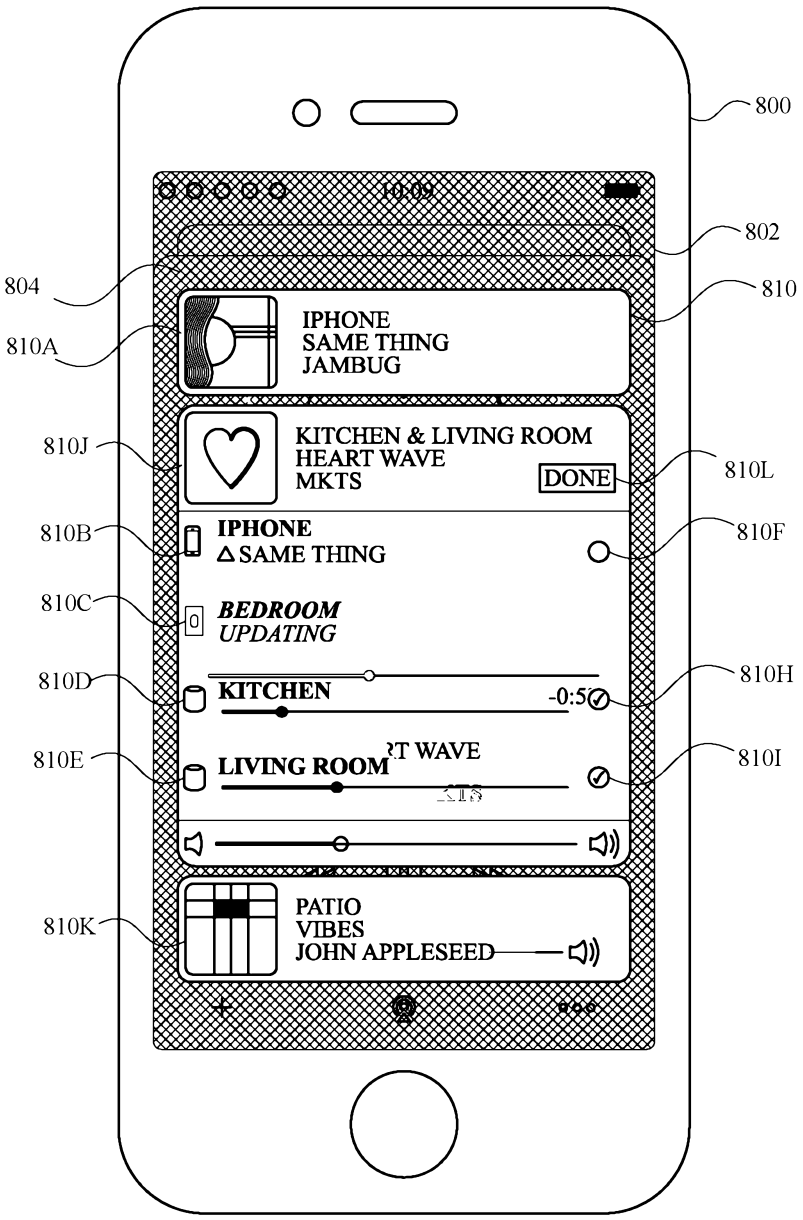

800
10:09
802
804
810
IPHONE
SAME THING
JAMBUG
810A
810J
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
DONE
810L
810B
IPHONE
△SAME THING
810F
810C
BEDROOM
UPDATING
810D
KITCHEN
-0:5
810H
810E
LIVING ROOM
810I
810K
PATIO
VIBES
JOHN APPLESEED

FIG. 8E

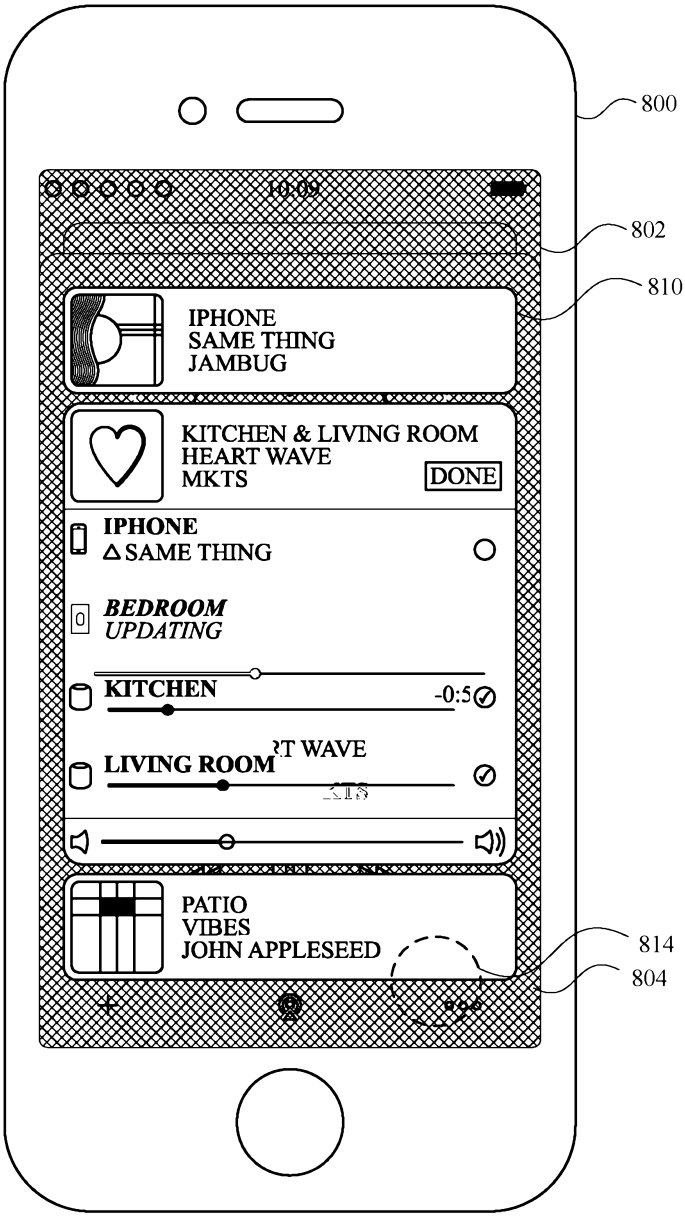

800
10:09
802
810
IPHONE
SAME THING
JAMBUG
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
DONE
IPHONE
SAME THING
BEDROOM
UPDATING
KITCHEN
-0:5
LIVING ROOM
PATIO
VIBES
JOHN APPLESEED
814
804

FIG. 8F

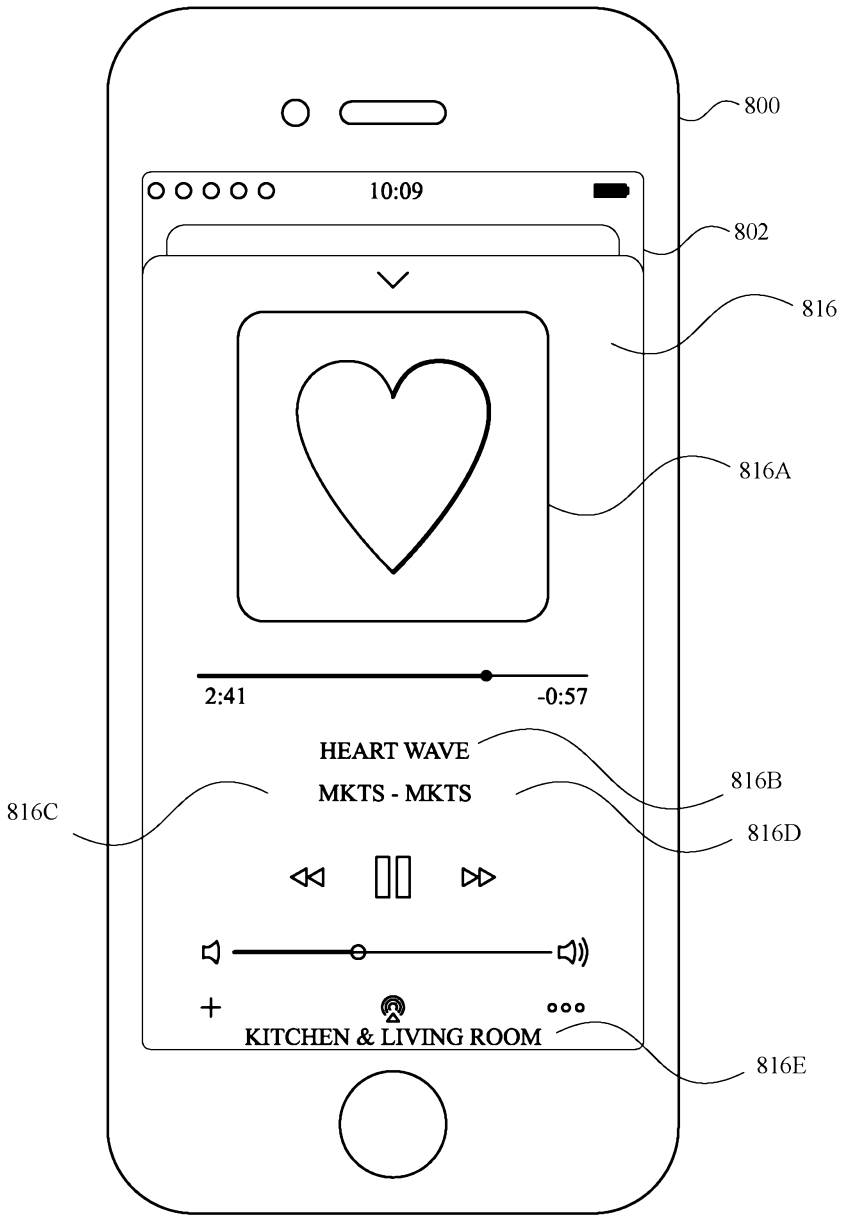

800
10:09
802
816
816A
2:41
-0:57
HEART WAVE
MKTS - MKTS
816B
816C
816D
KITCHEN & LIVING ROOM
816E

FIG. 8G

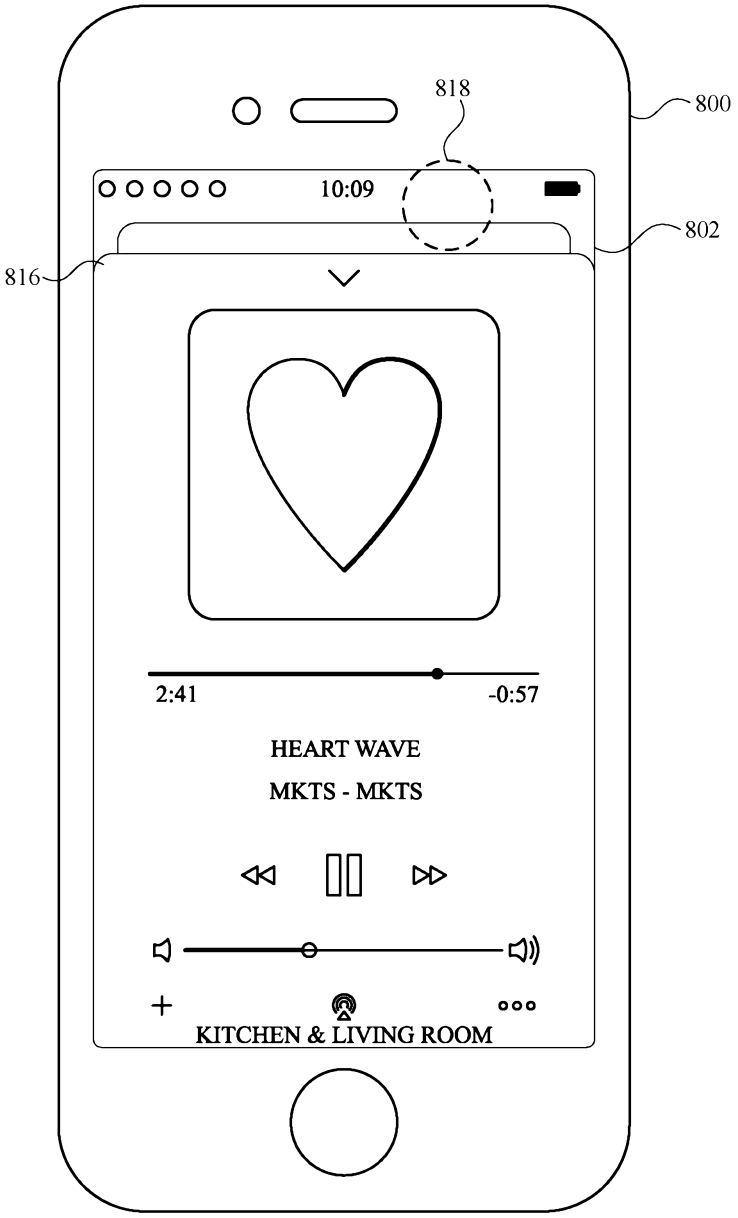

818
800
10:09
802
816
2:41
-0:57
HEART WAVE
MKTS - MKTS
KITCHEN & LIVING ROOM

FIG. 8H

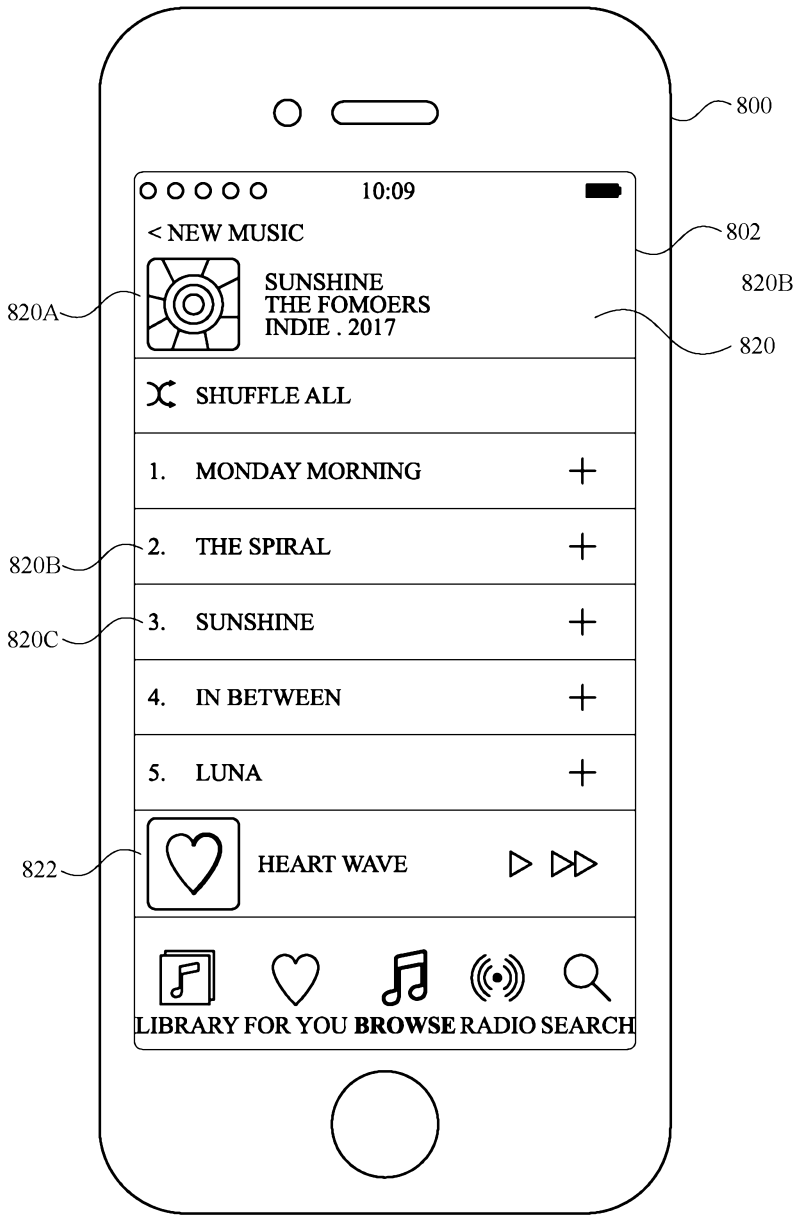

800
10:09
< NEW MUSIC
802
820B
820A
SUNSHINE
THE FOMOERS
INDIE . 2017
820
SHUFFLE ALL
1. MONDAY MORNING
820B
2. THE SPIRAL
820C
3. SUNSHINE
4. IN BETWEEN
5. LUNA
822
HEART WAVE
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 8I

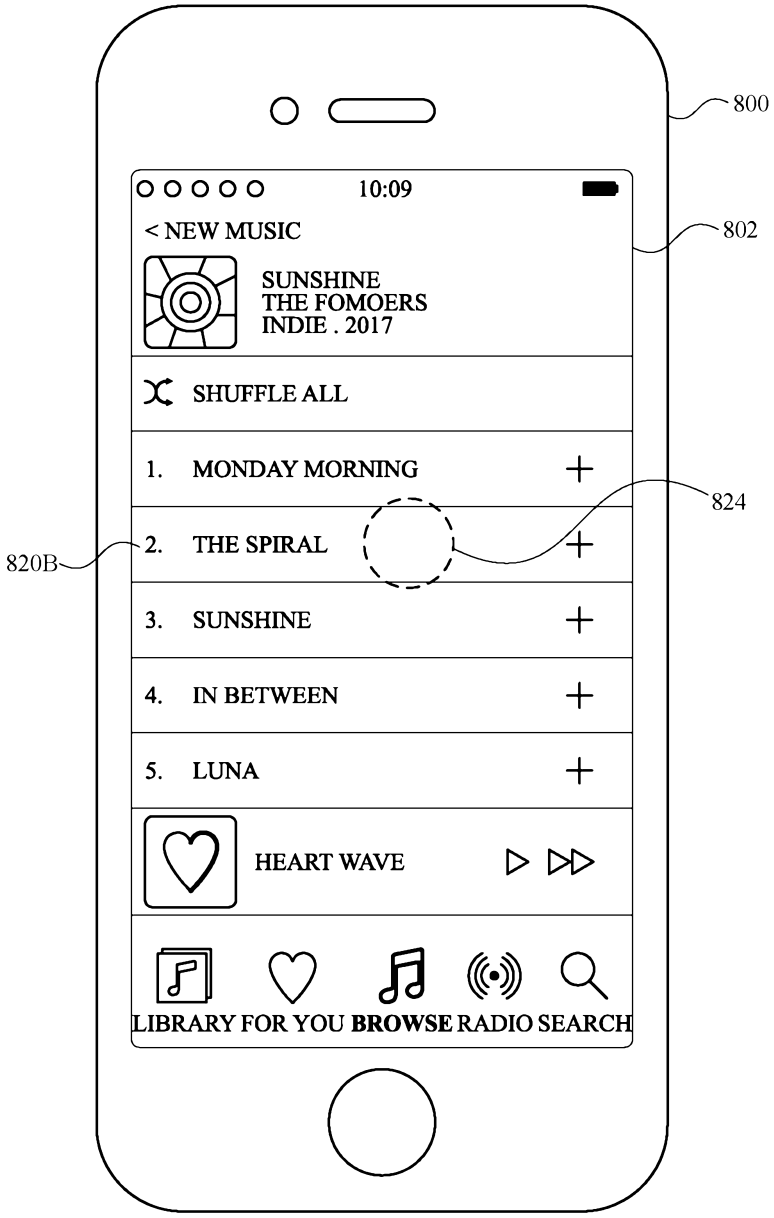

800
10:09
< NEW MUSIC
802
SUNSHINE
THE FOMOERS
INDIE . 2017
SHUFFLE ALL
1. MONDAY MORNING
824
820B
2. THE SPIRAL
3. SUNSHINE
4. IN BETWEEN
5. LUNA
HEART WAVE
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 8J

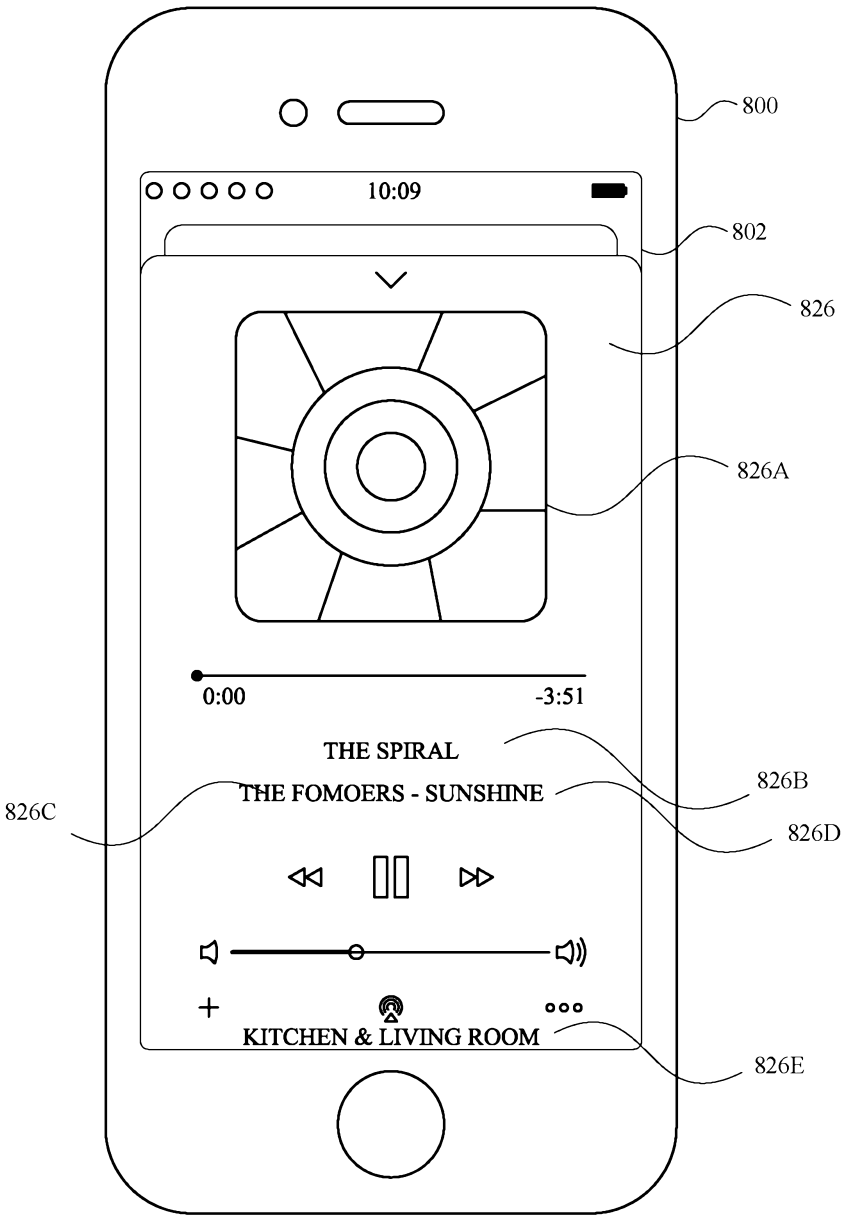

800
10:09
802
826
826A
0:00
-3:51
THE SPIRAL
826B
826C
THE FOMOERS - SUNSHINE
826D
KITCHEN & LIVING ROOM
826E

FIG. 8K

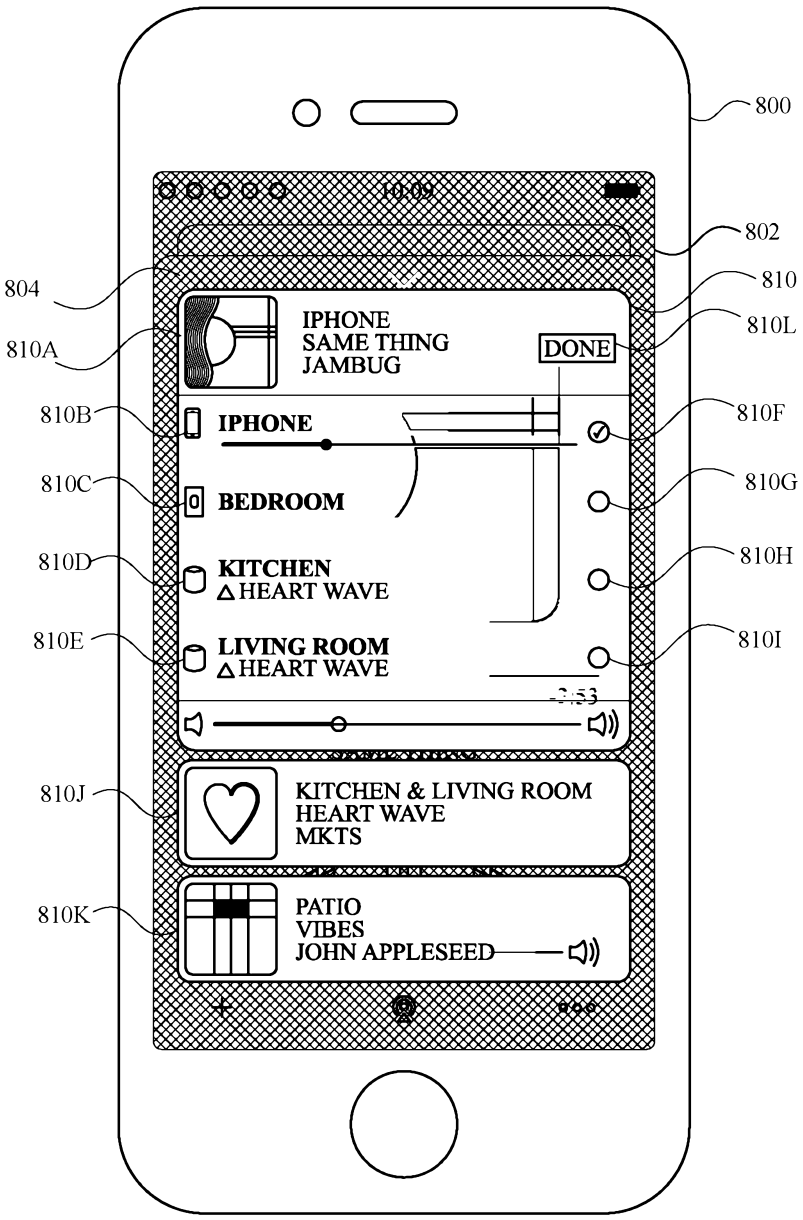

800
10:09
802
804
810
IPHONE
SAME THING
JAMBUG
DONE
810L
810A
810B
IPHONE
810F
810C
BEDROOM
810G
810D
KITCHEN
△HEART WAVE
810H
810E
LIVING ROOM
△HEART WAVE
810I
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
810J
PATIO
VIBES
JOHN APPLESEED
810K

FIG. 8L

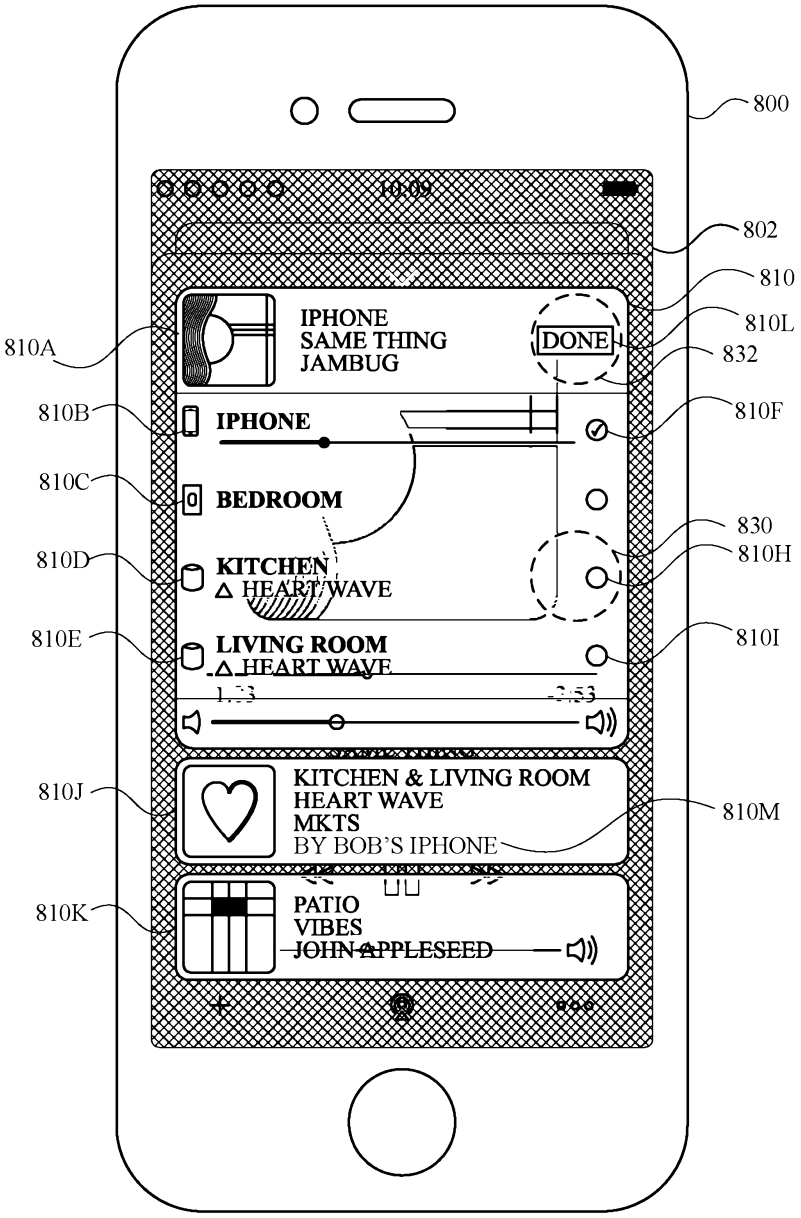

800
10:09
802
810
IPHONE
SAME THING
JAMBUG
DONE
810L
832
810A
810B
IPHONE
810F
810C
BEDROOM
830
810D
KITCHEN
HEART WAVE
810H
810E
LIVING ROOM
HEART WAVE
810I
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
BY BOB'S IPHONE
810J
810M
PATIO
VIBES
JOHN APPLESEED
810K

FIG. 8M

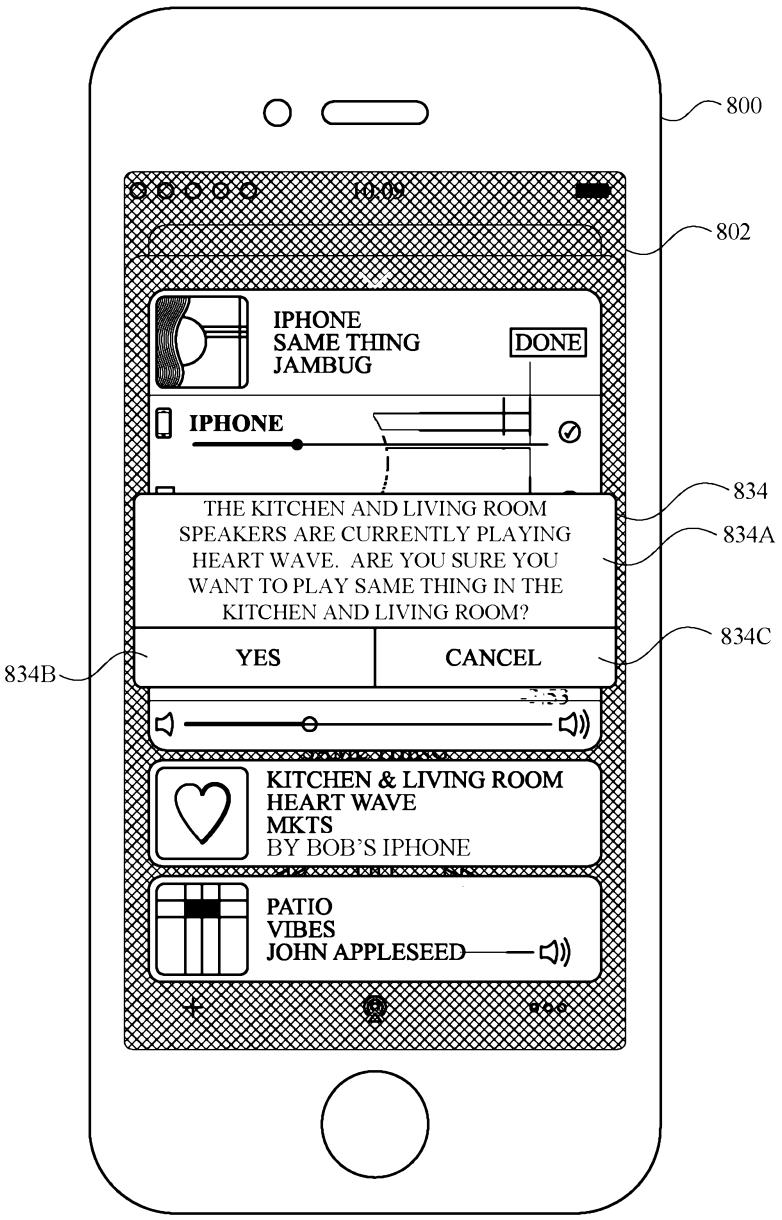

800
10:09
802
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
834
THE KITCHEN AND LIVING ROOM SPEAKERS ARE CURRENTLY PLAYING HEART WAVE. ARE YOU SURE YOU WANT TO PLAY SAME THING IN THE KITCHEN AND LIVING ROOM?
834A
834B
YES
CANCEL
834C
KITCHEN & LIVING ROOM
HEART WAVE
MKTS
BY BOB'S IPHONE
PATIO
VIBES
JOHN APPLESEED

FIG. 8N

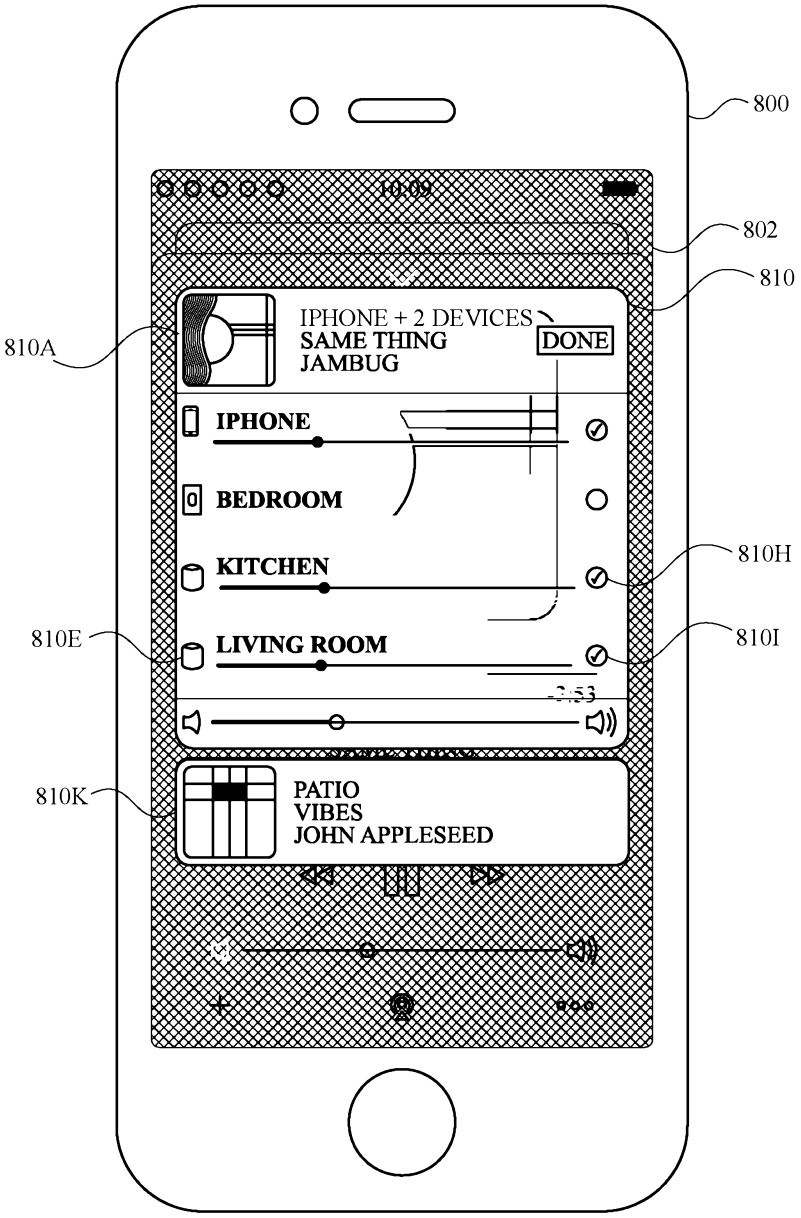

800
10:09
802
810
IPHONE + 2 DEVICES
SAME THING
JAMBUG
DONE
810A
IPHONE
BEDROOM
KITCHEN
LIVING ROOM
810H
810E
810I
PATIO
VIBES
JOHN APPLESEED
810K

FIG. 80

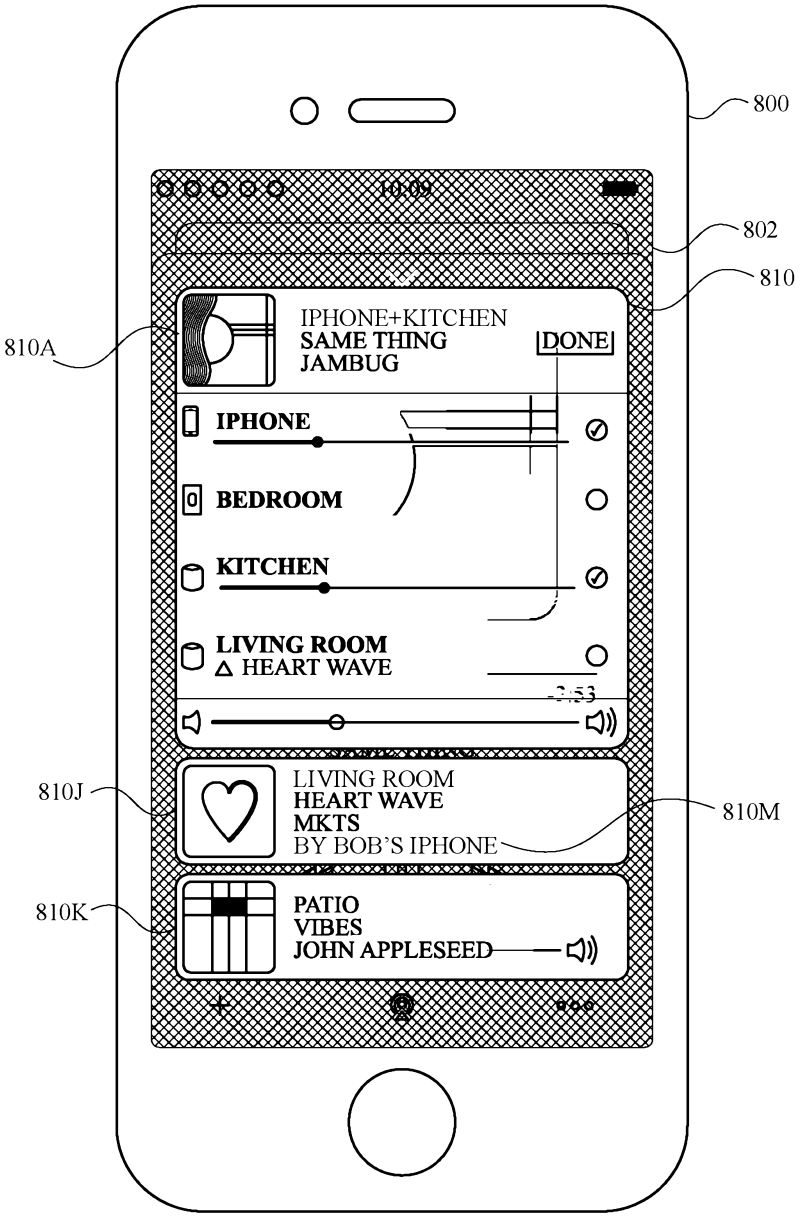

800
10:09
802
810
IPHONE+KITCHEN
SAME THING
JAMBUG
DONE
810A
IPHONE
BEDROOM
KITCHEN
LIVING ROOM
△ HEART WAVE
LIVING ROOM
HEART WAVE
MKTS
BY BOB'S IPHONE
810J
810M
PATIO
VIBES
JOHN APPLESEED
810K

FIG. 8P

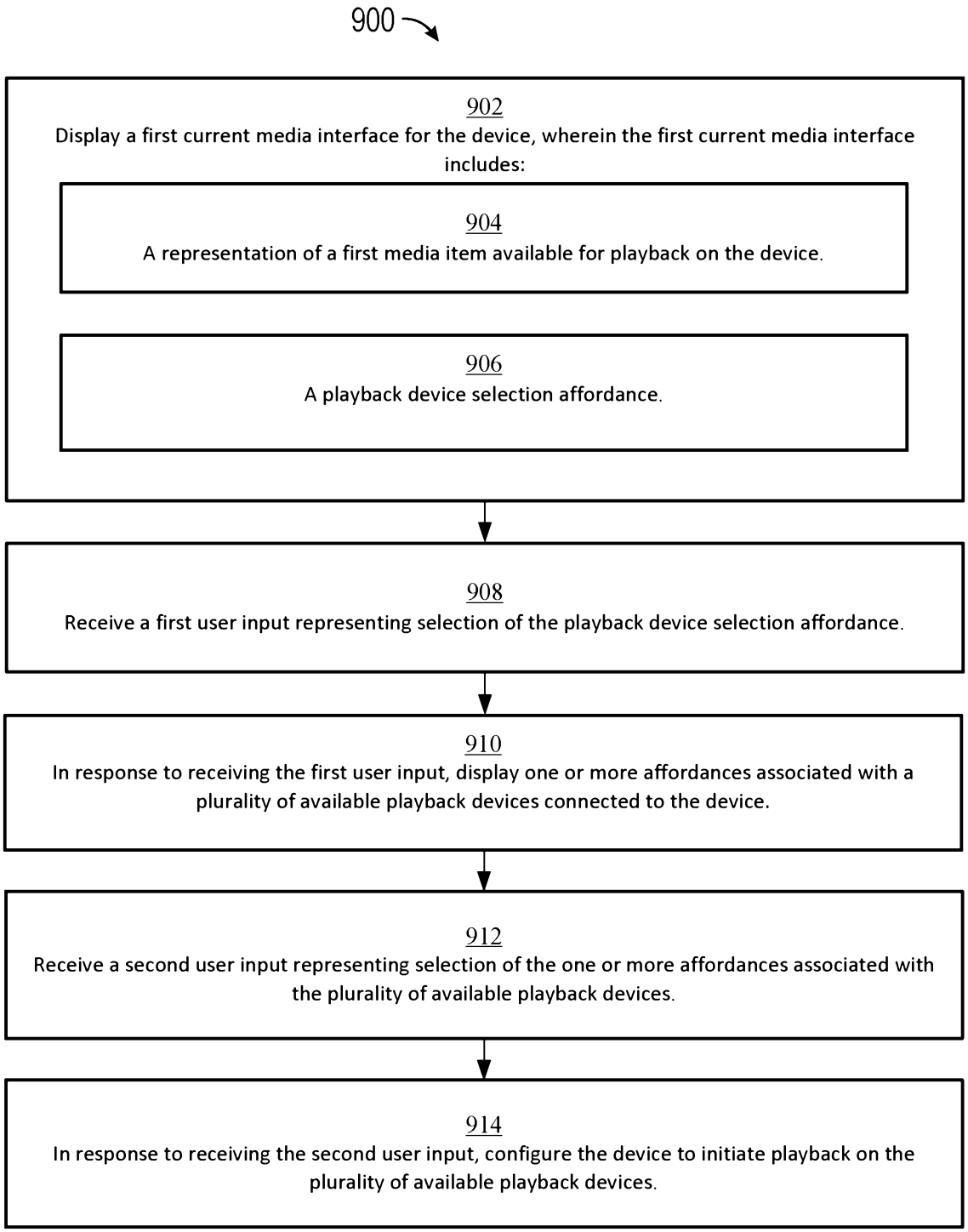

900
902
Display a first current media interface for the device, wherein the first current media interface includes:
904
A representation of a first media item available for playback on the device.
906
A playback device selection affordance.
908
Receive a first user input representing selection of the playback device selection affordance.
910
In response to receiving the first user input, display one or more affordances associated with a plurality of available playback devices connected to the device.
912
Receive a second user input representing selection of the one or more affordances associated with the plurality of available playback devices.
914
In response to receiving the second user input, configure the device to initiate playback on the plurality of available playback devices.

FIG. 9

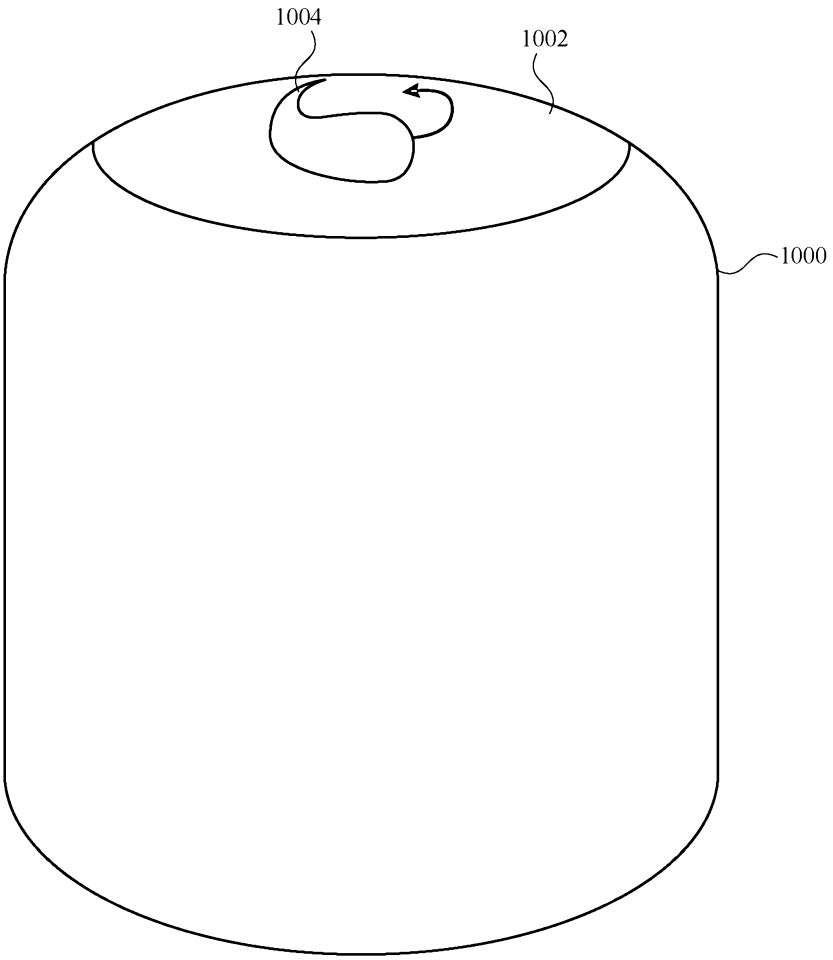


FIG. 10A

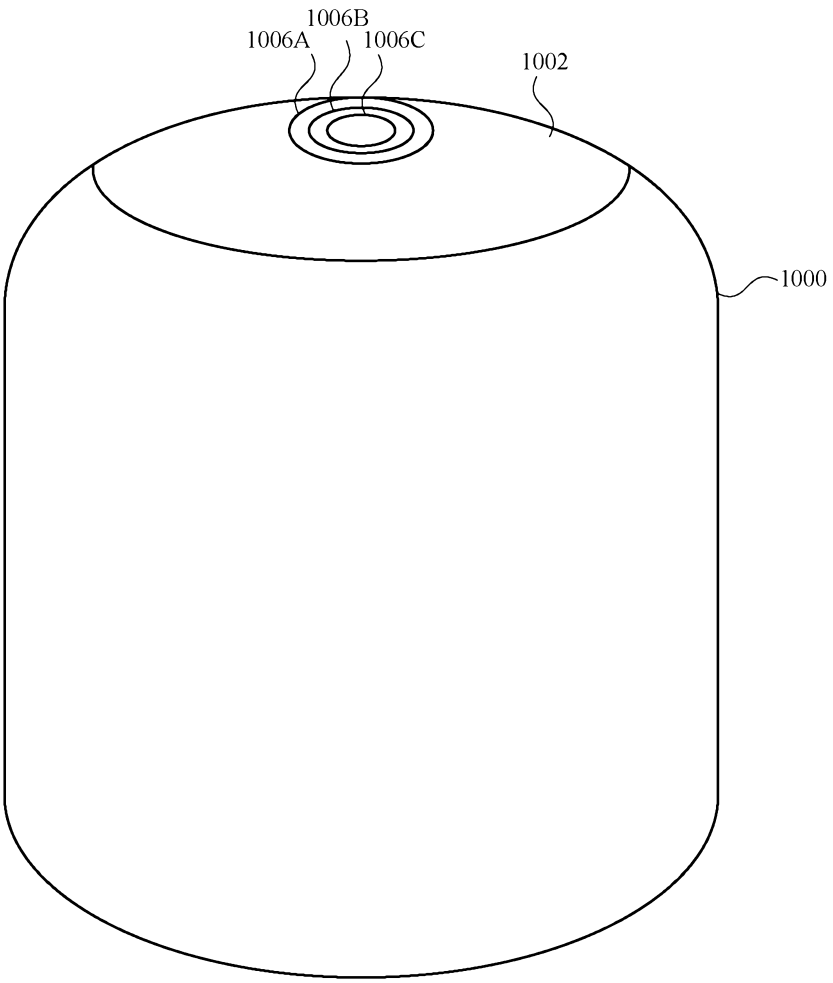


FIG. 10B

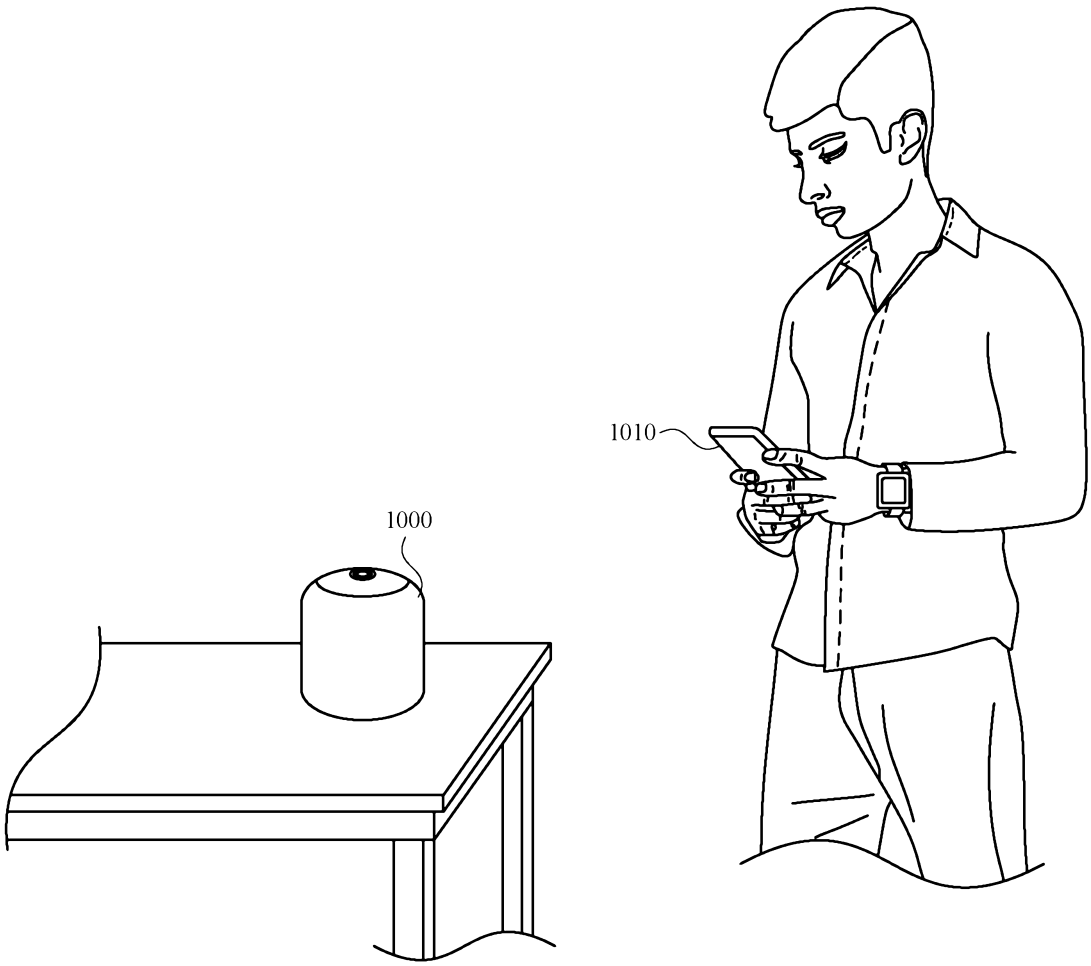


FIG. 10C

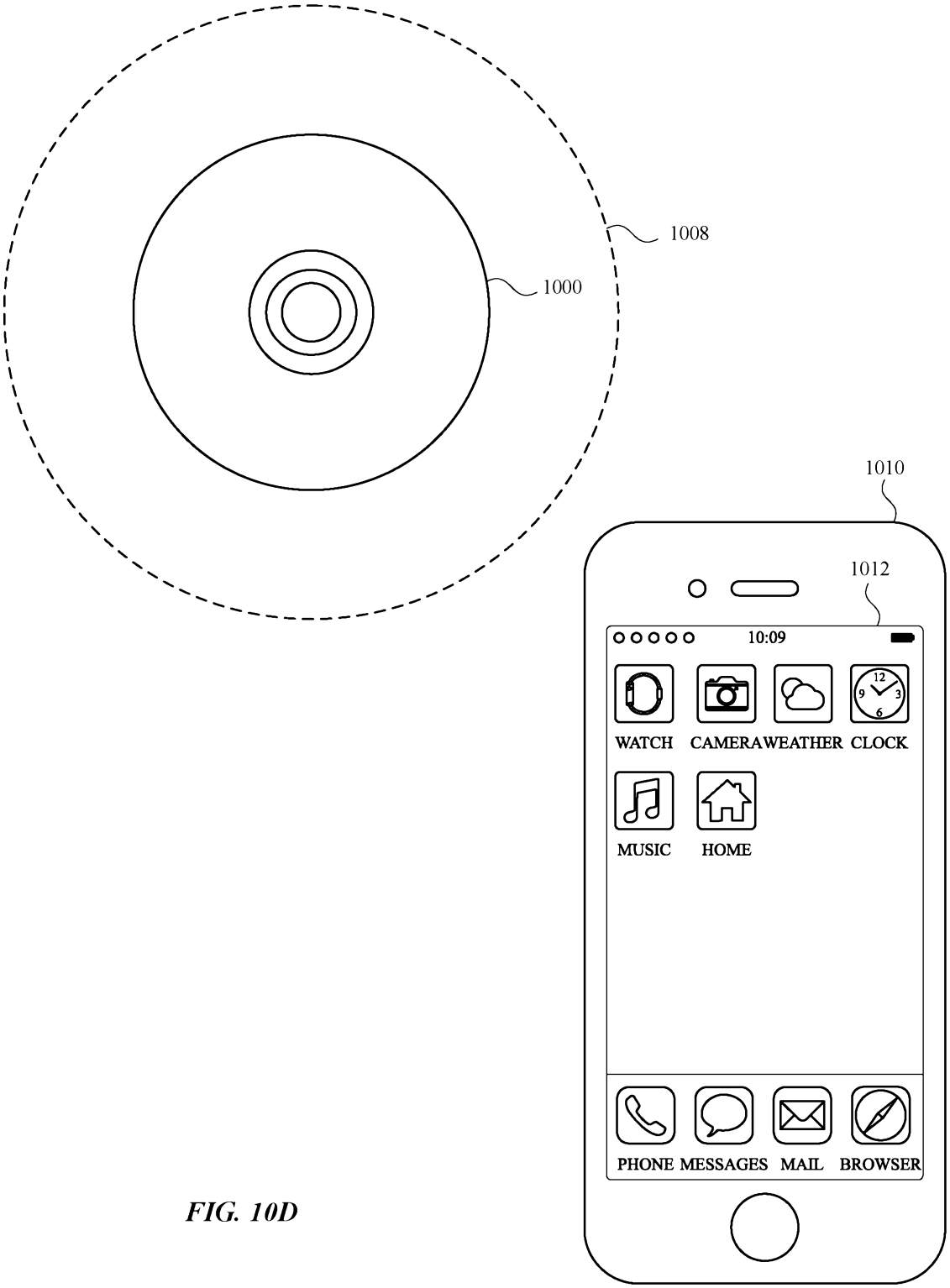

1008
1000
1010
1012
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
PHONE
MESSAGES
MAIL
BROWSER

FIG. 10D

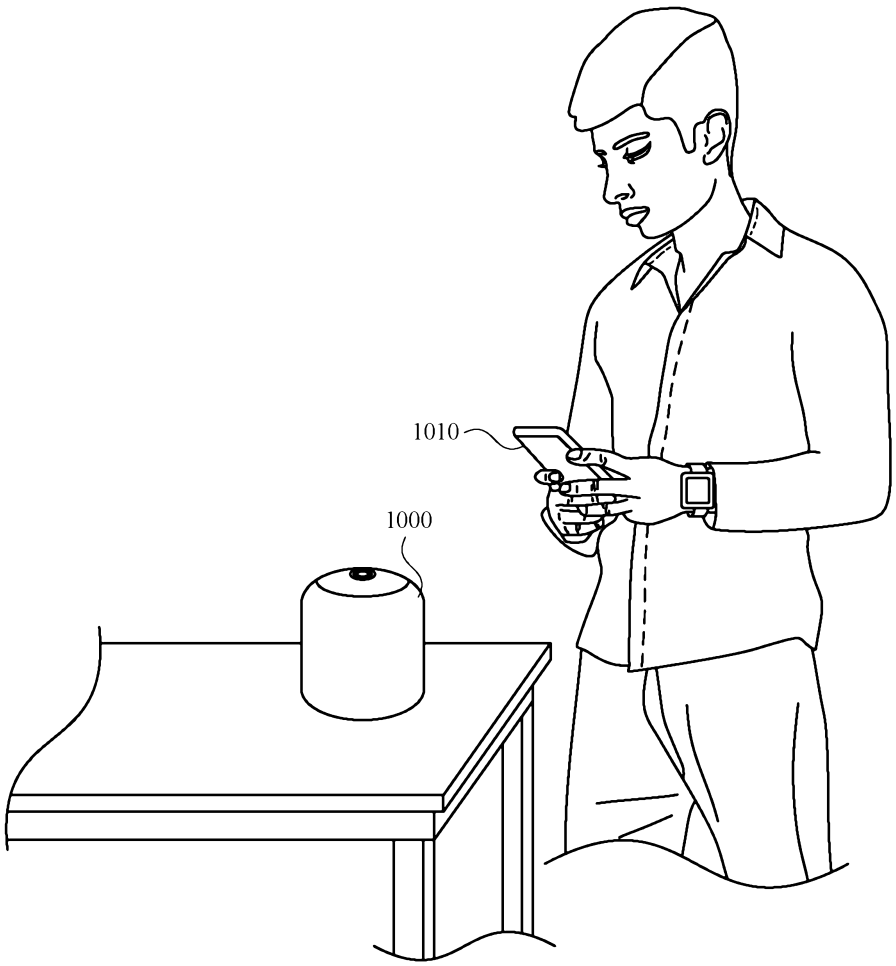


FIG. 10E

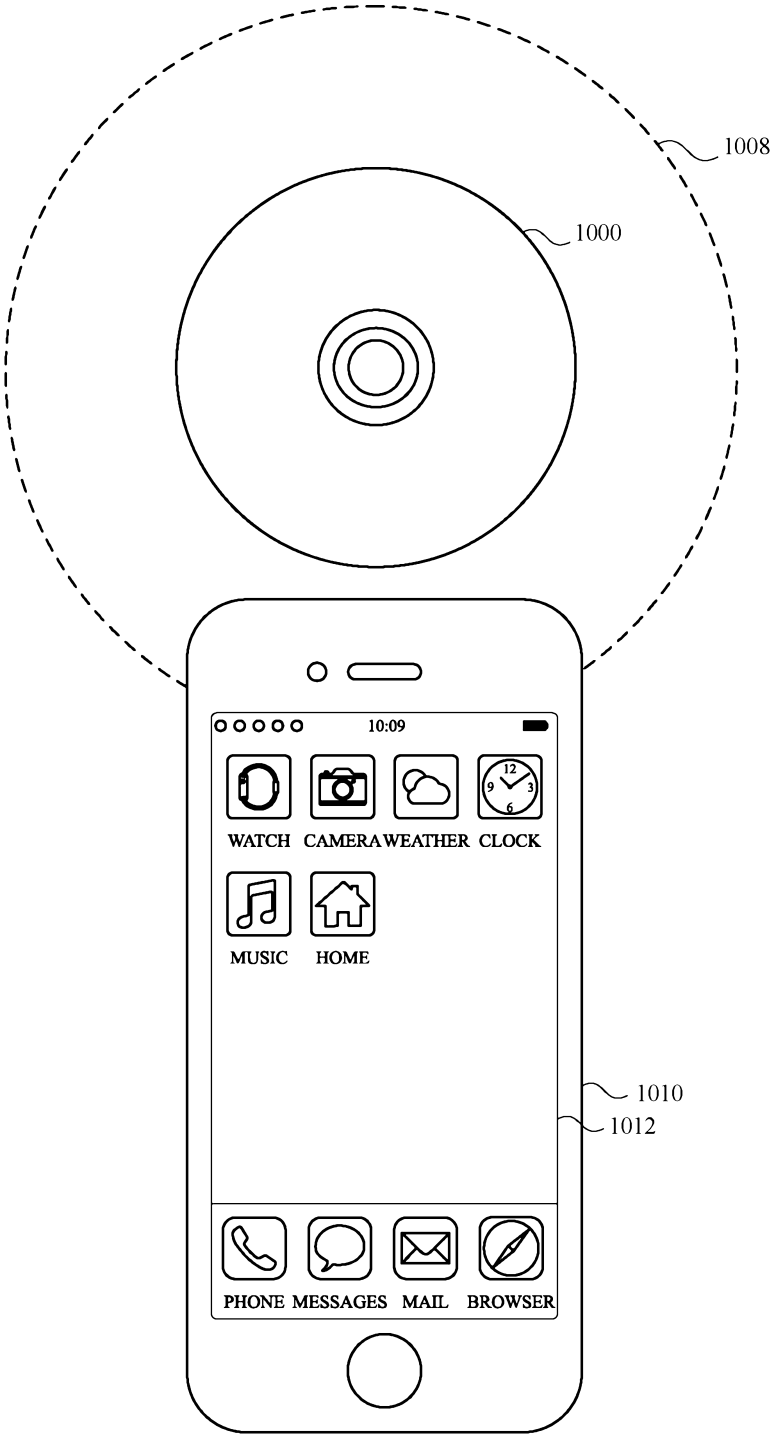

1008
1000
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1010
1012
PHONE
MESSAGES
MAIL
BROWSER

FIG. 10F

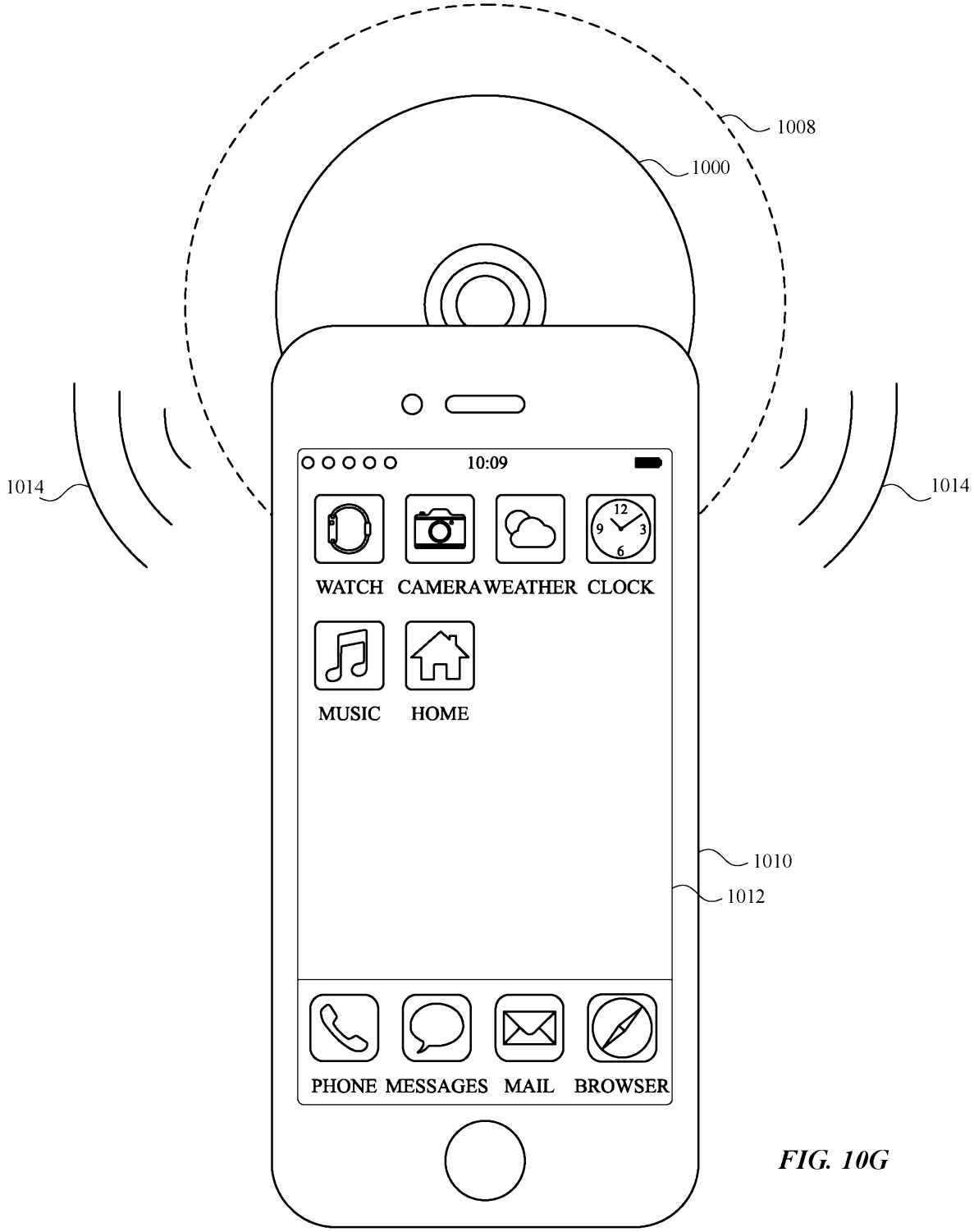

1008
1000
1014
1014
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1010
1012
PHONE
MESSAGES
MAIL
BROWSER

FIG. 10G

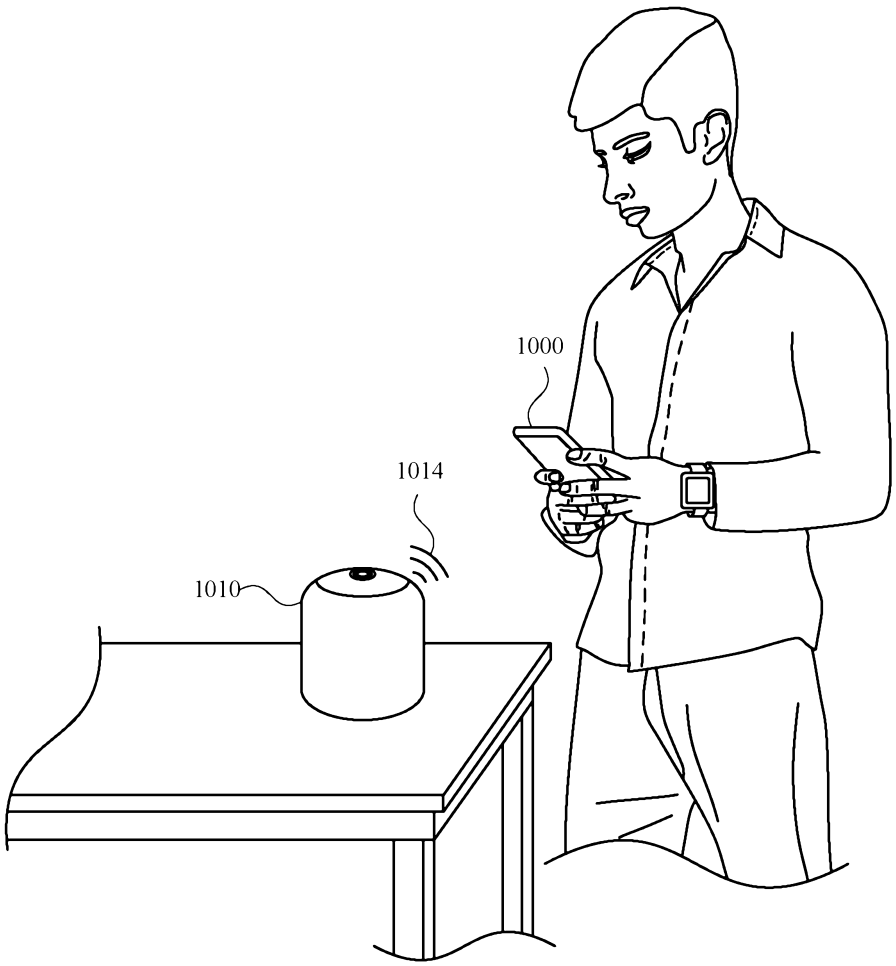


FIG. 10H

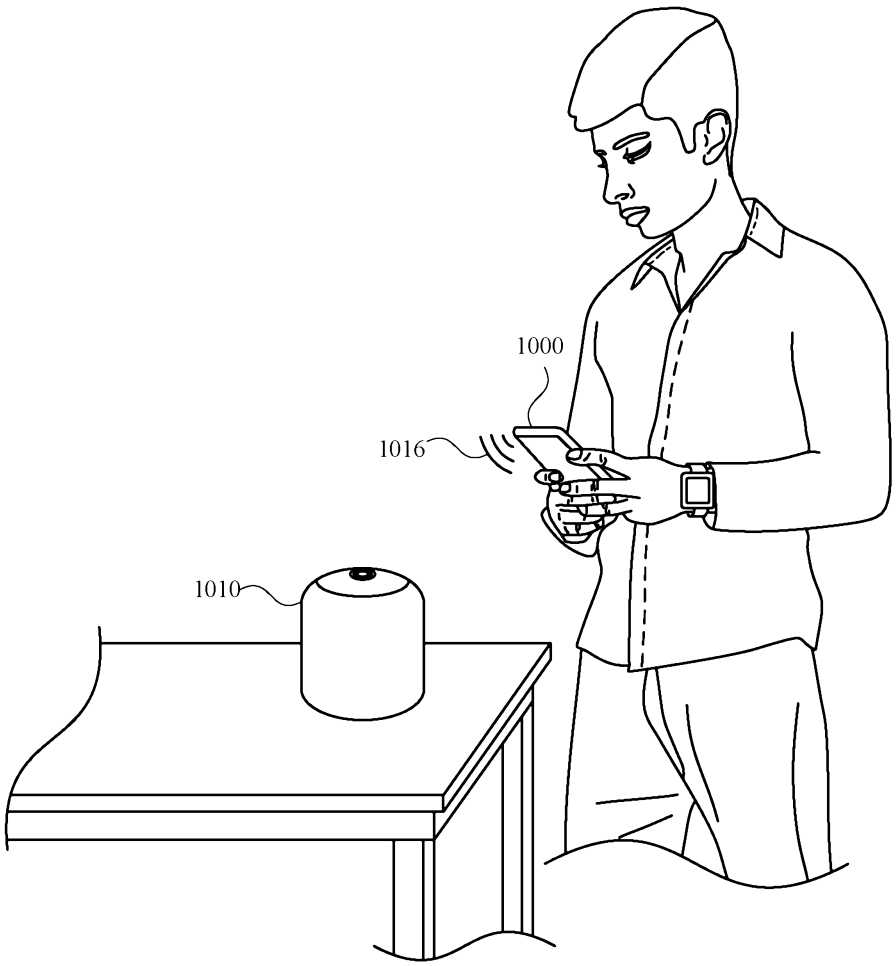


FIG. 10I

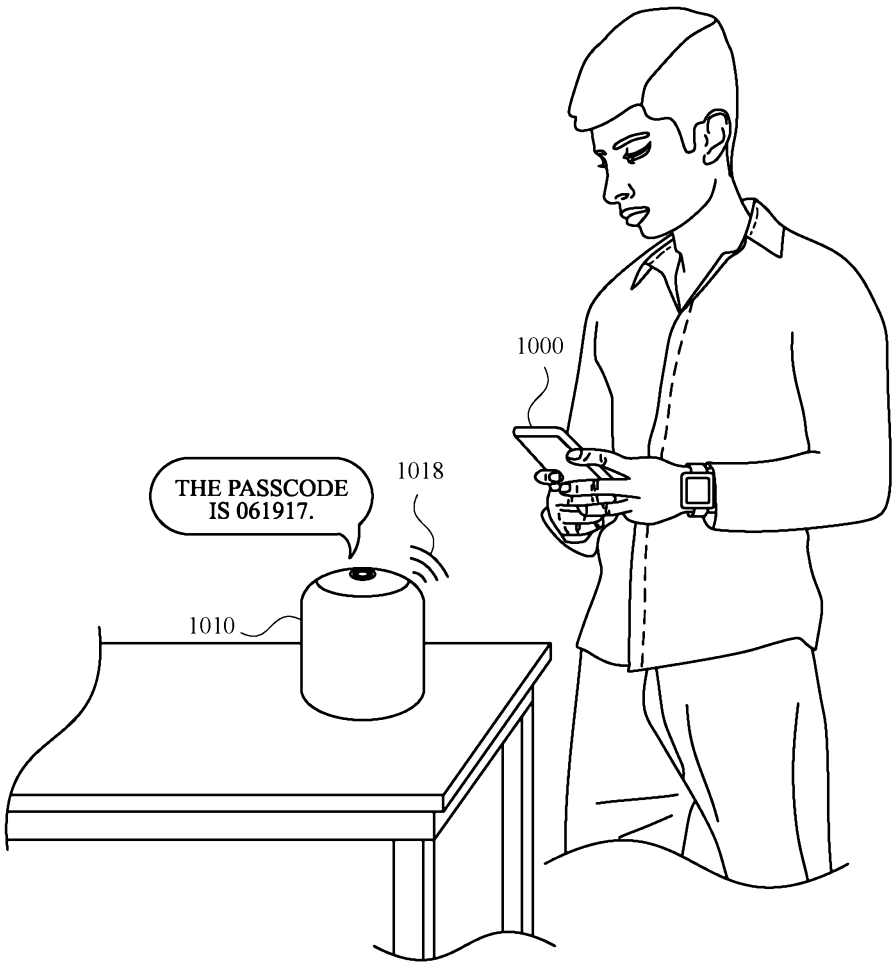


FIG. 10J

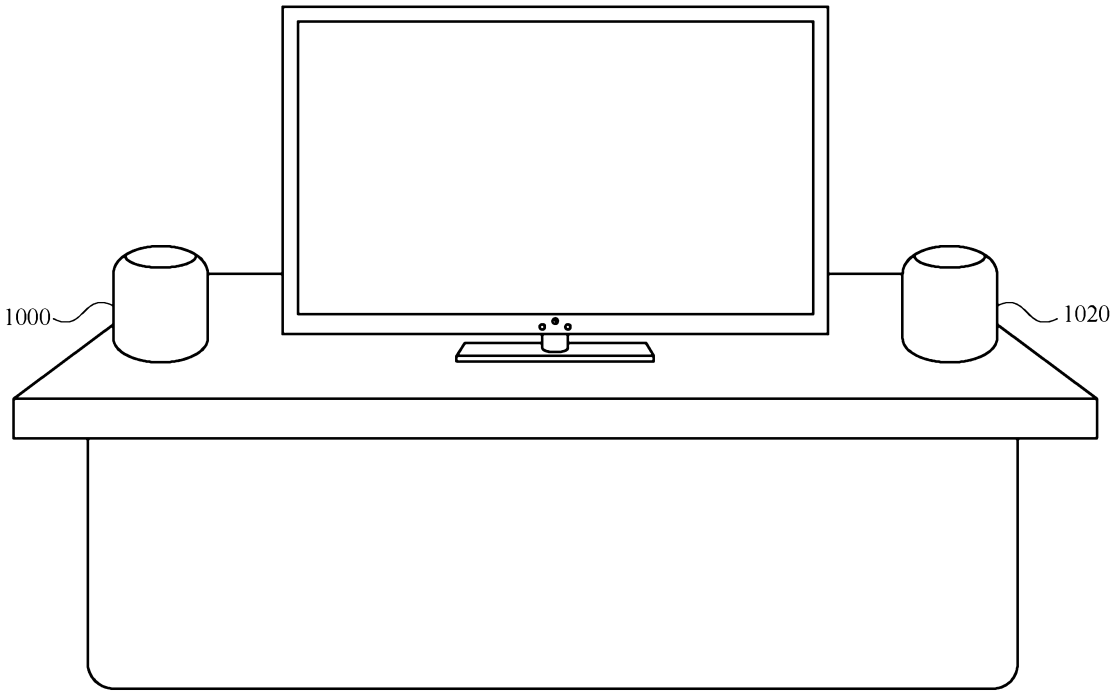


FIG. 10K

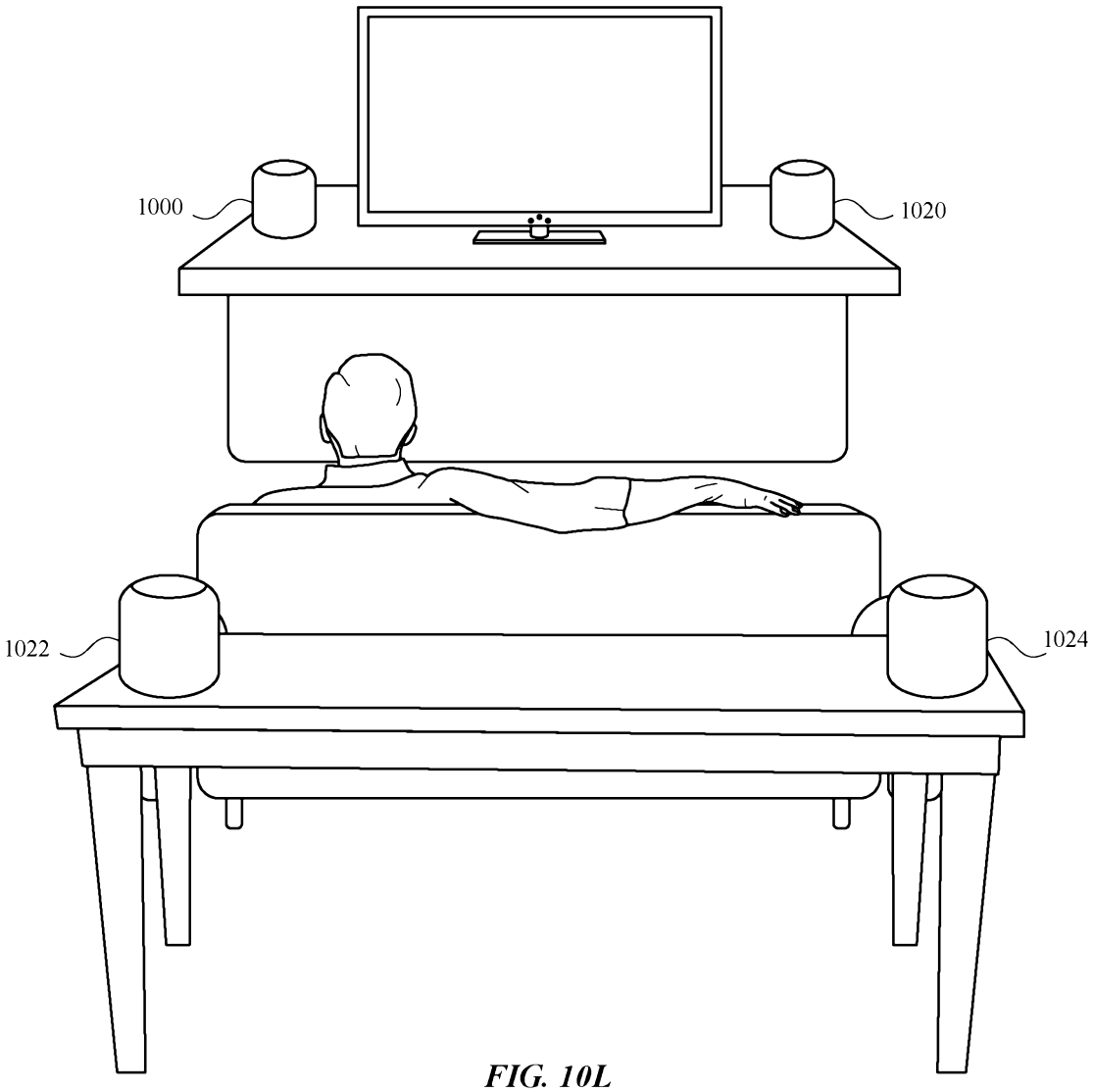


FIG. 10L

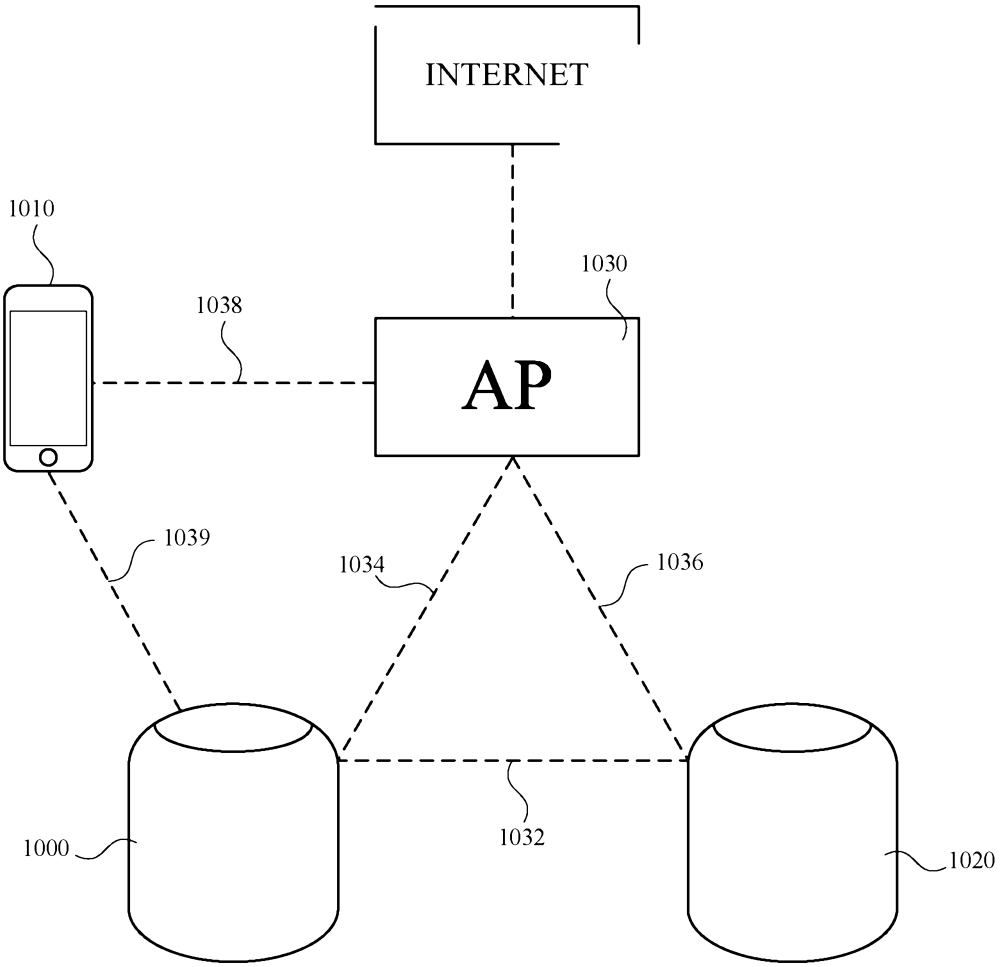


FIG. 10M

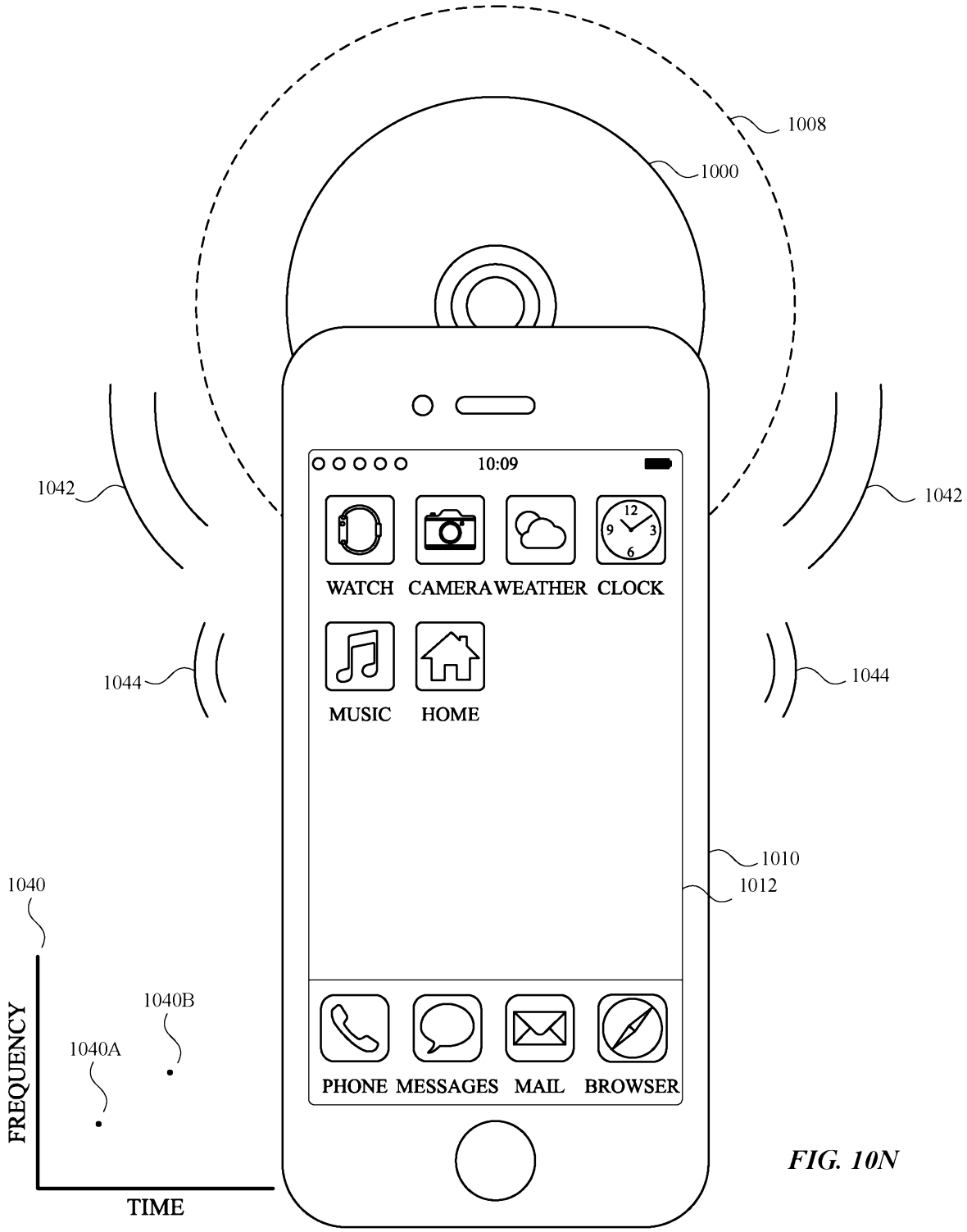

1008
1000
1042
1042
10:09
WATCH
CAMERA
WEATHER
CLOCK
1044
1044
MUSIC
HOME
1010
1012
1040
1040B
1040A
FREQUENCY
TIME
PHONE
MESSAGES
MAIL
BROWSER

FIG. 10N

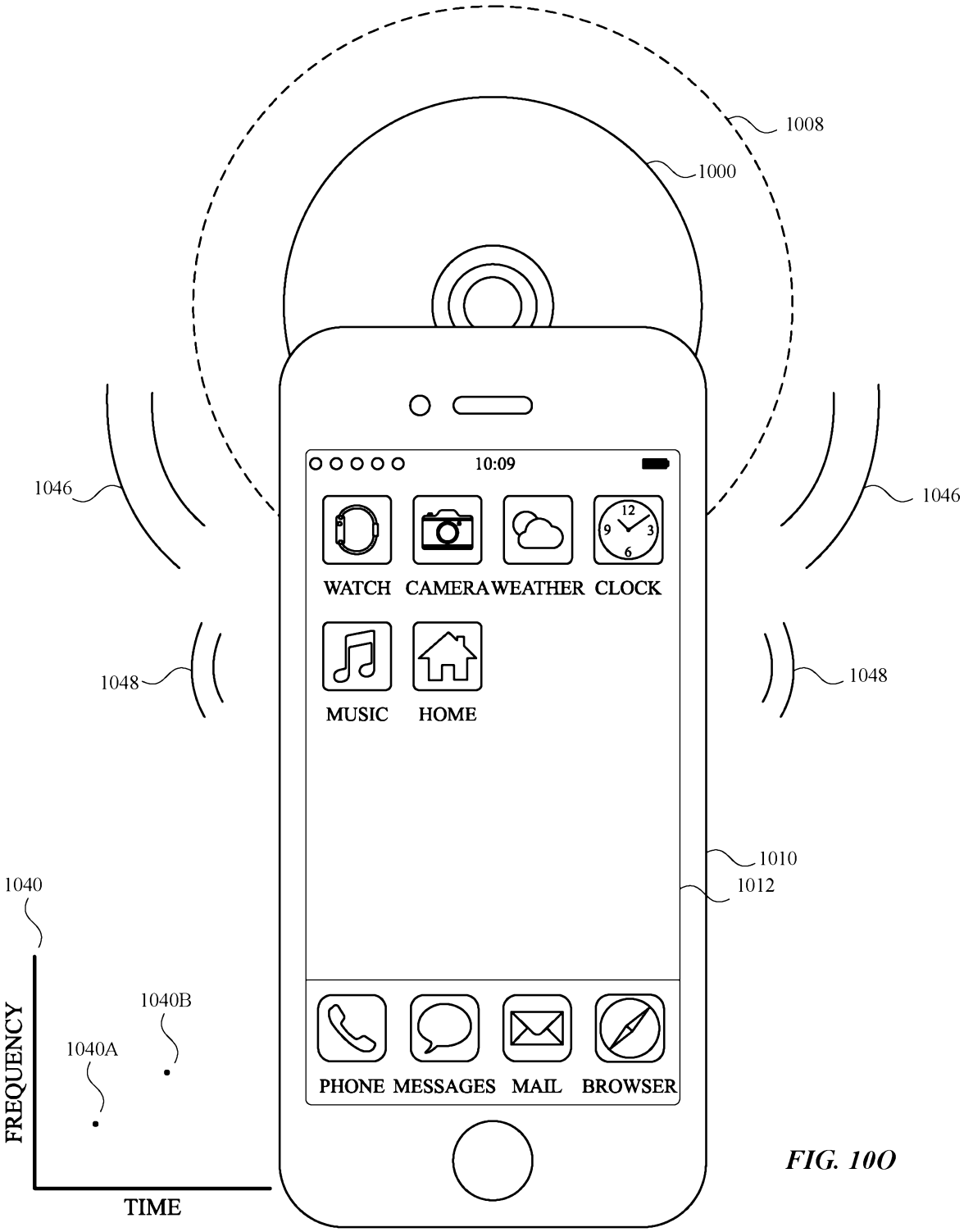

1008
1000
1046
1046
10:09
WATCH
CAMERA
WEATHER
CLOCK
1048
1048
MUSIC
HOME
1010
1012
1040
1040B
1040A
FREQUENCY
TIME
PHONE
MESSAGES
MAIL
BROWSER

FIG. 10O

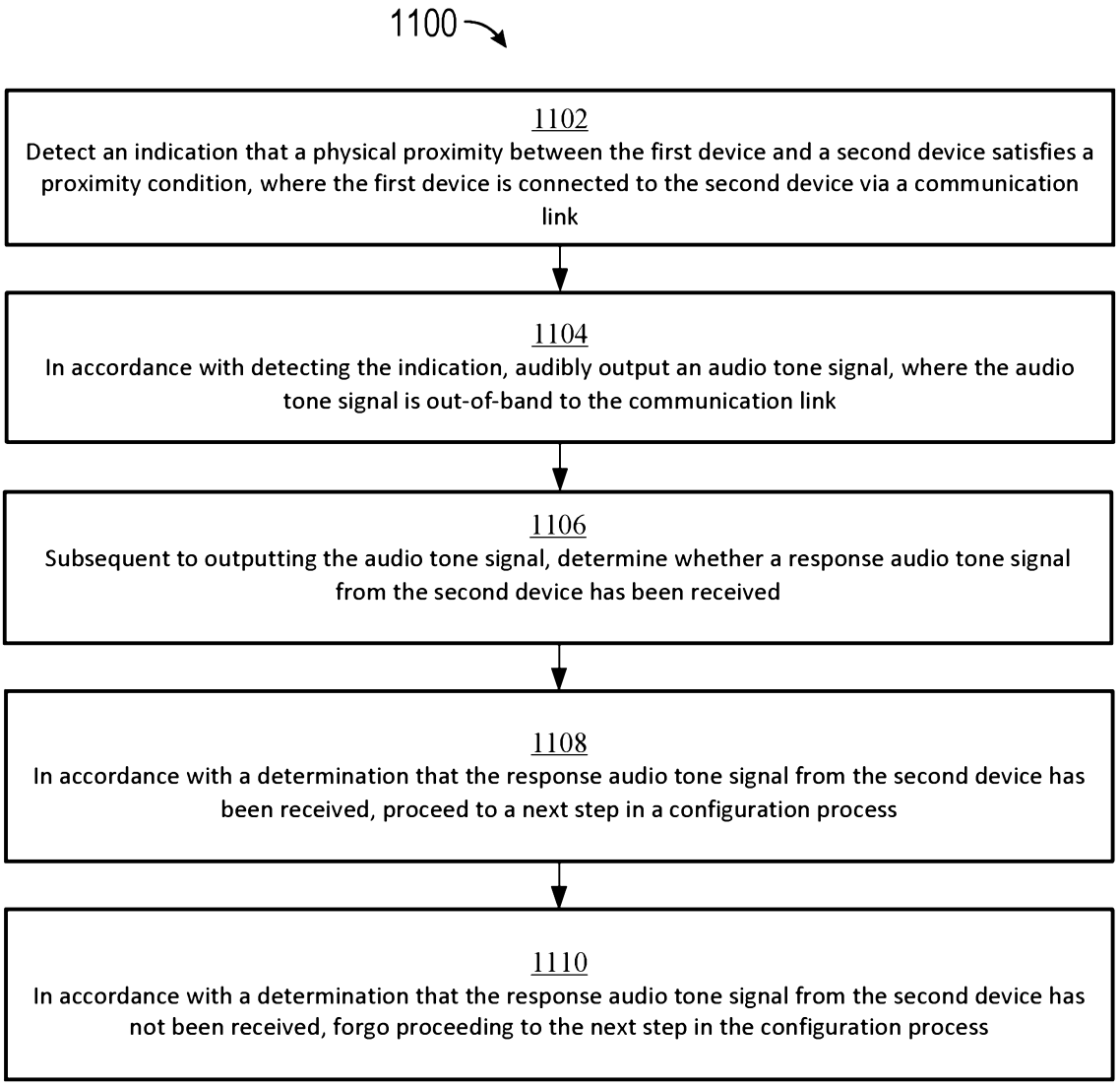

1100
1102
Detect an indication that a physical proximity between the first device and a second device satisfies a proximity condition, where the first device is connected to the second device via a communication link
1104
In accordance with detecting the indication, audibly output an audio tone signal, where the audio tone signal is out-of-band to the communication link
1106
Subsequent to outputting the audio tone signal, determine whether a response audio tone signal from the second device has been received
1108
In accordance with a determination that the response audio tone signal from the second device has been received, proceed to a next step in a configuration process
1110
In accordance with a determination that the response audio tone signal from the second device has not been received, forgo proceeding to the next step in the configuration process

FIG. 11

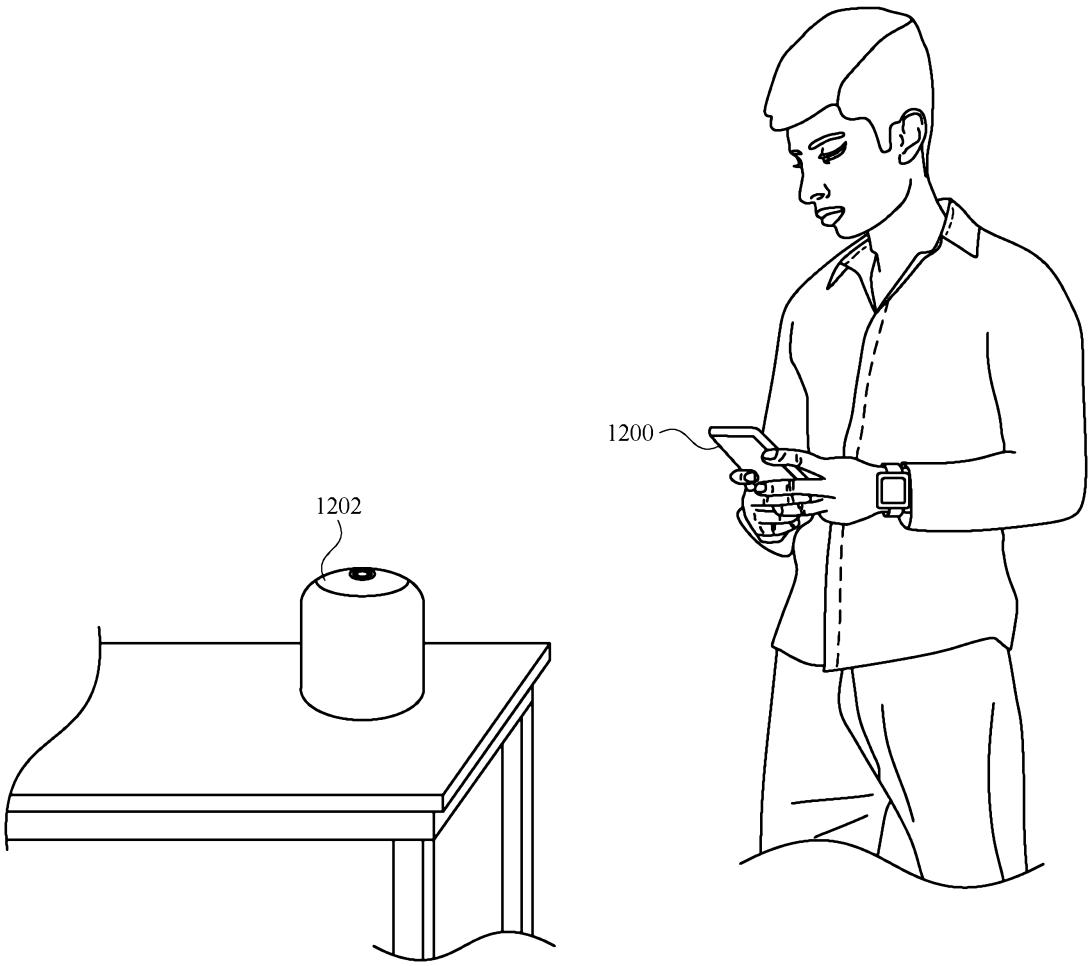


FIG. 12A

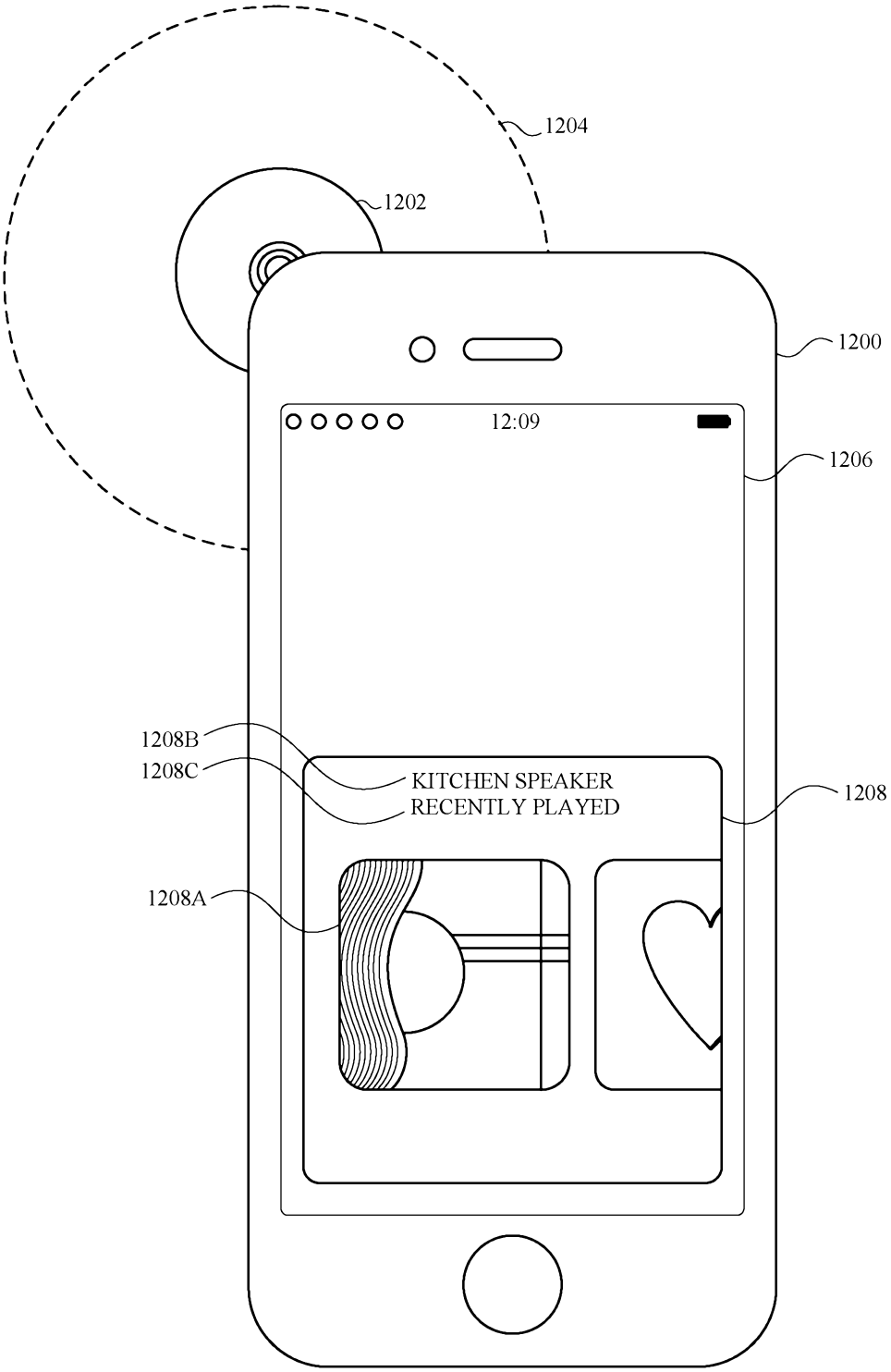

1204
1202
1200
12:09
1206
1208B
1208C
KITCHEN SPEAKER
RECENTLY PLAYED
1208
1208A

FIG. 12B

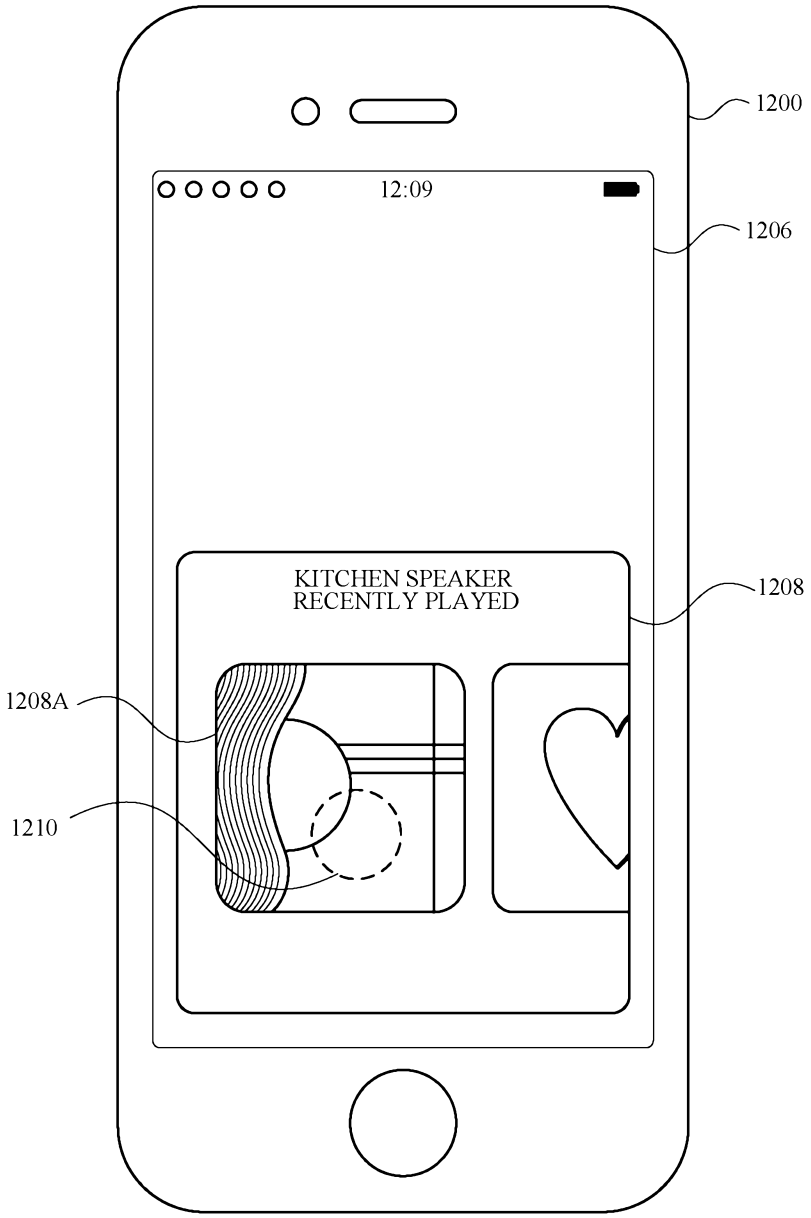

1200
12:09
1206
KITCHEN SPEAKER
RECENTLY PLAYED
1208
1208A
1210

FIG. 12C

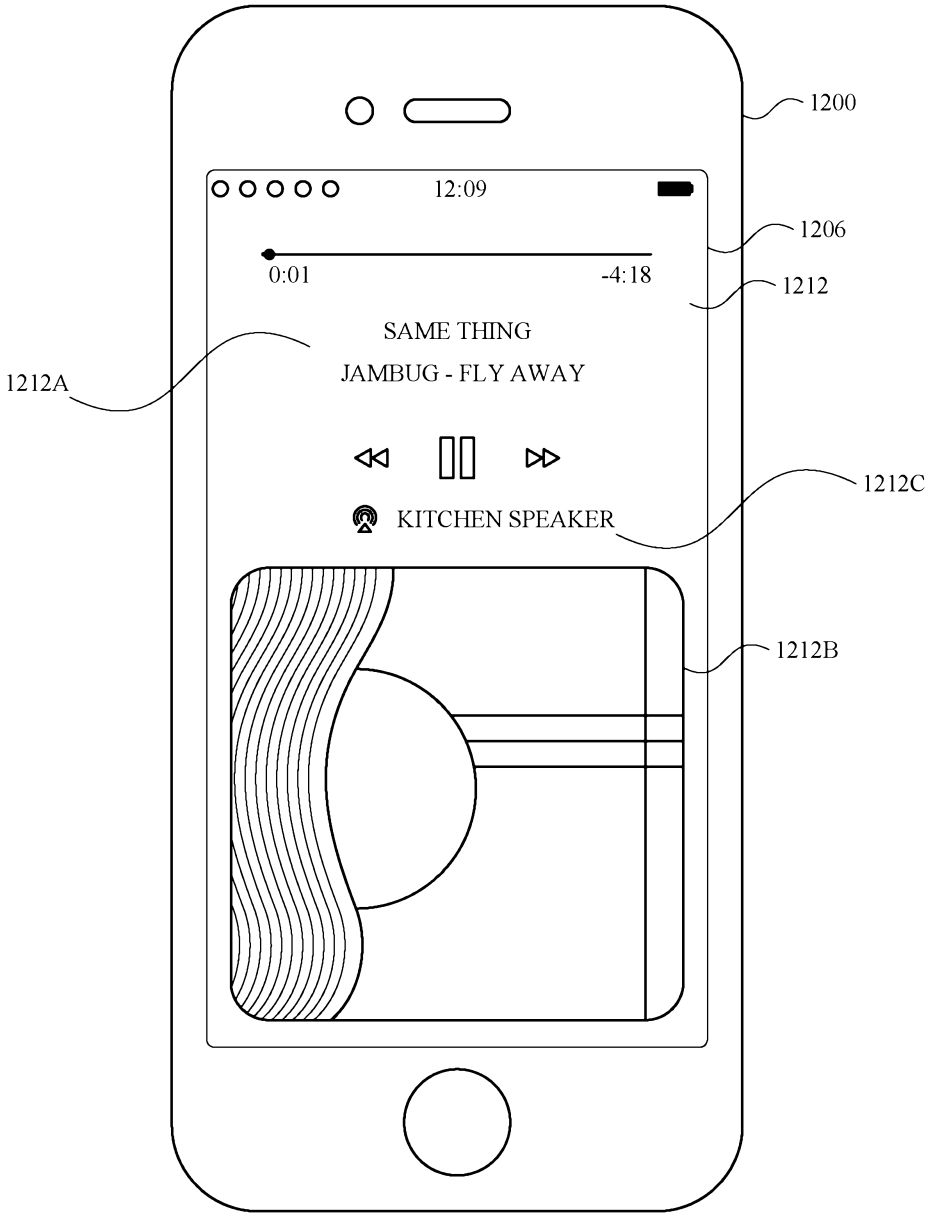

1200
12:09
1206
0:01
-4:18
1212
SAME THING
JAMBUG - FLY AWAY
1212A
1212C
KITCHEN SPEAKER
1212B

FIG. 12D

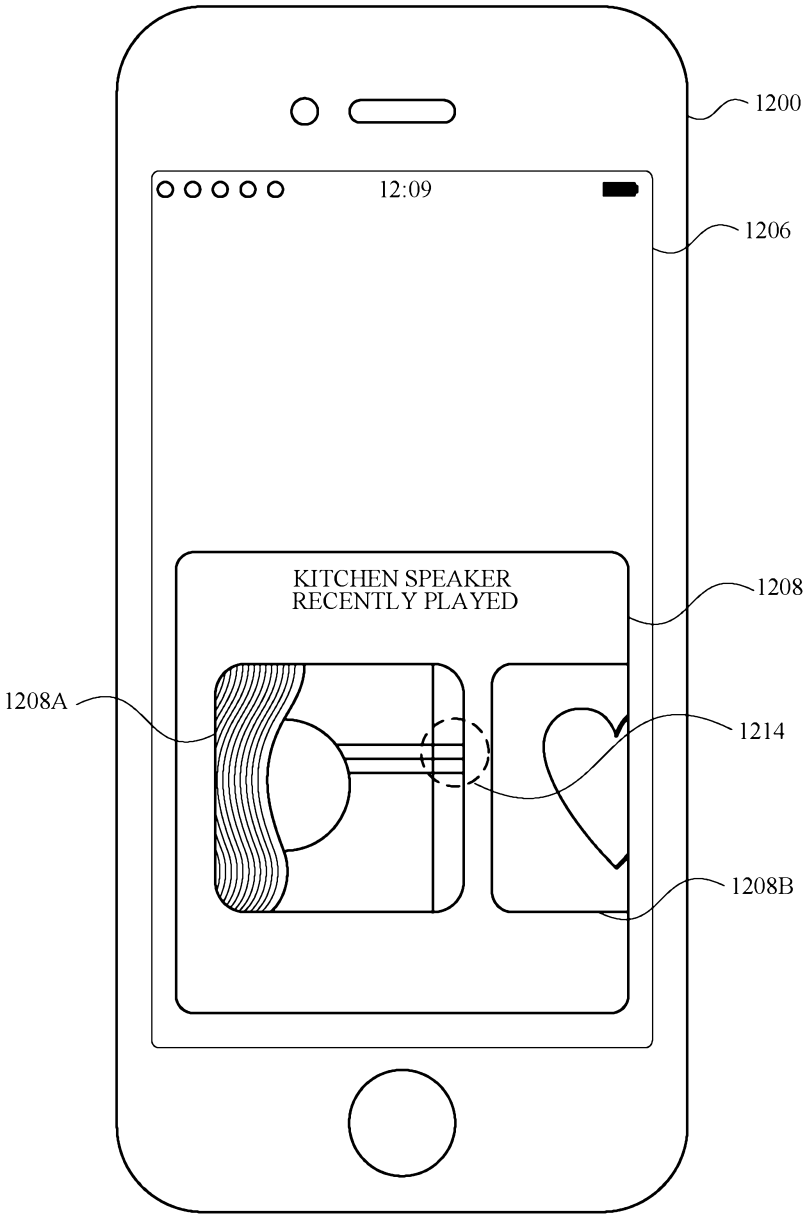

1200
12:09
1206
KITCHEN SPEAKER
RECENTLY PLAYED
1208
1208A
1214
1208B

FIG. 12E

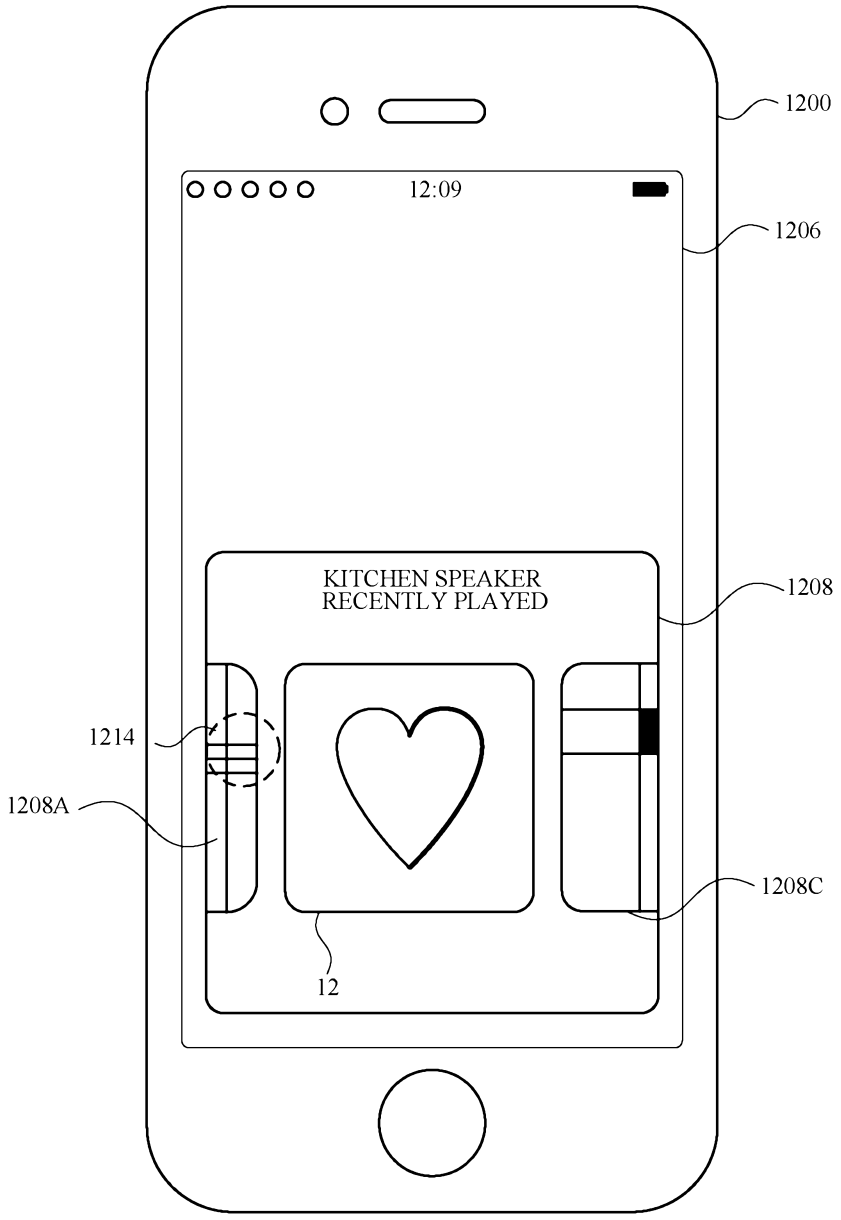

1200
12:09
1206
KITCHEN SPEAKER
RECENTLY PLAYED
1208
1214
1208A
1208C
12

FIG. 12F

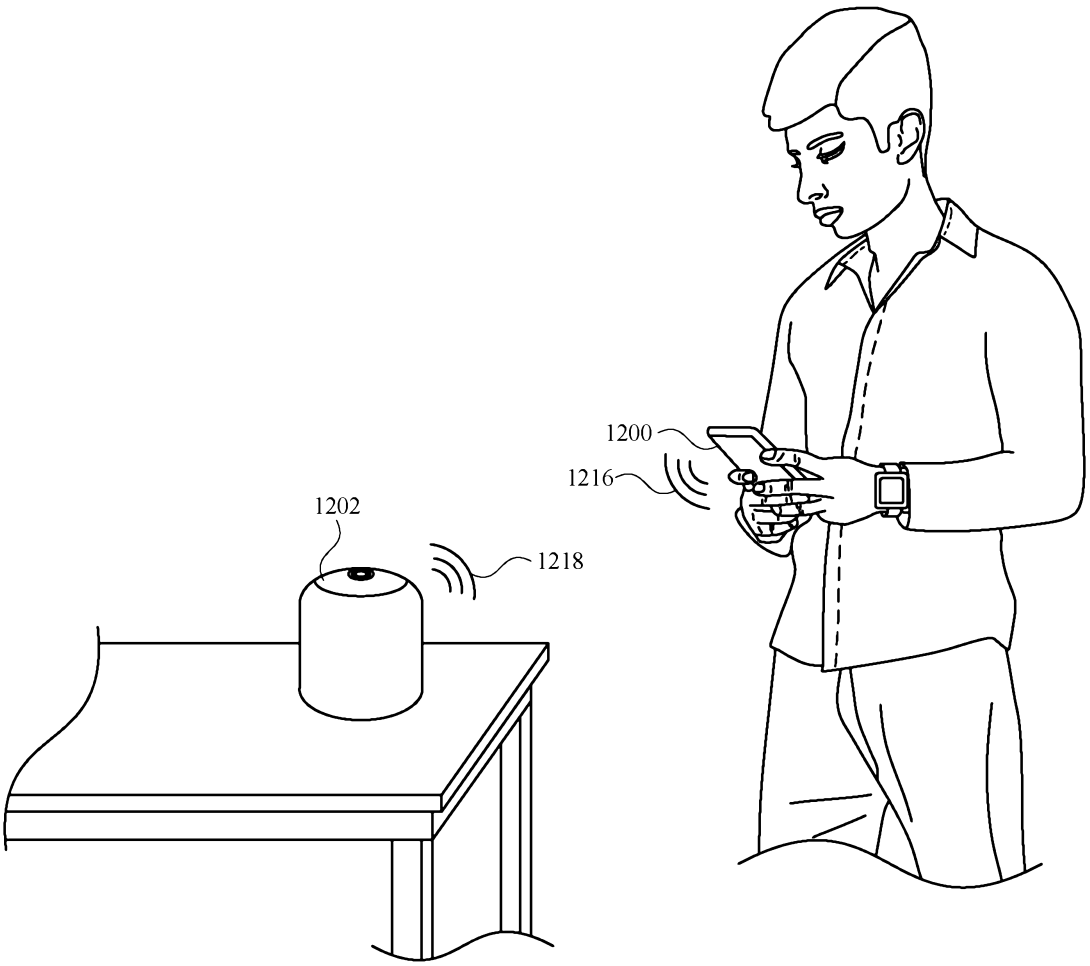


FIG. 12G

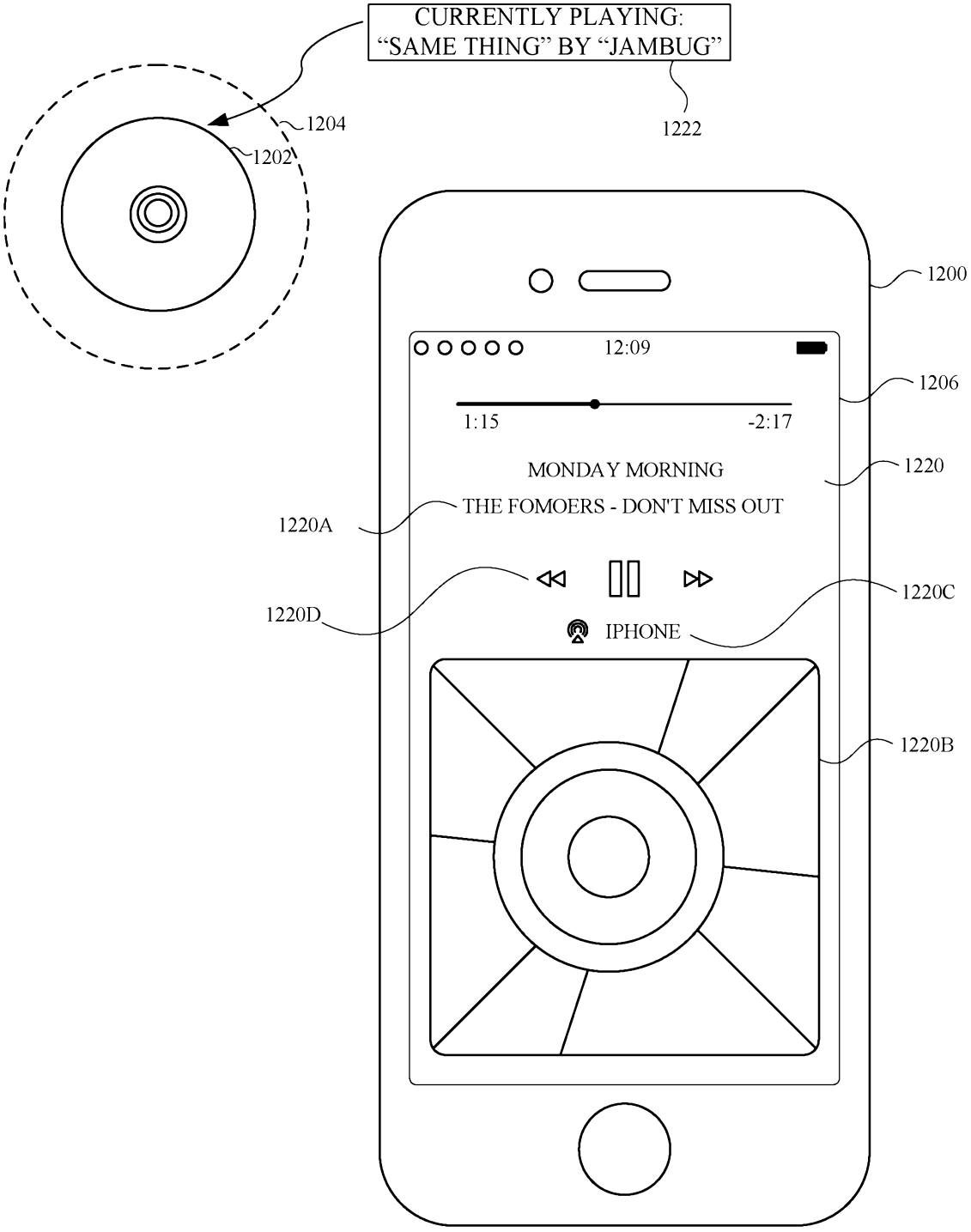

CURRENTLY PLAYING:
"SAME THING" BY "JAMBUG"
1222
1204
1202
1200
12:09
1206
1:15
-2:17
MONDAY MORNING
1220
THE FOMOERS - DON'T MISS OUT
1220A
1220D
1220C
IPHONE
1220B

FIG. 12H

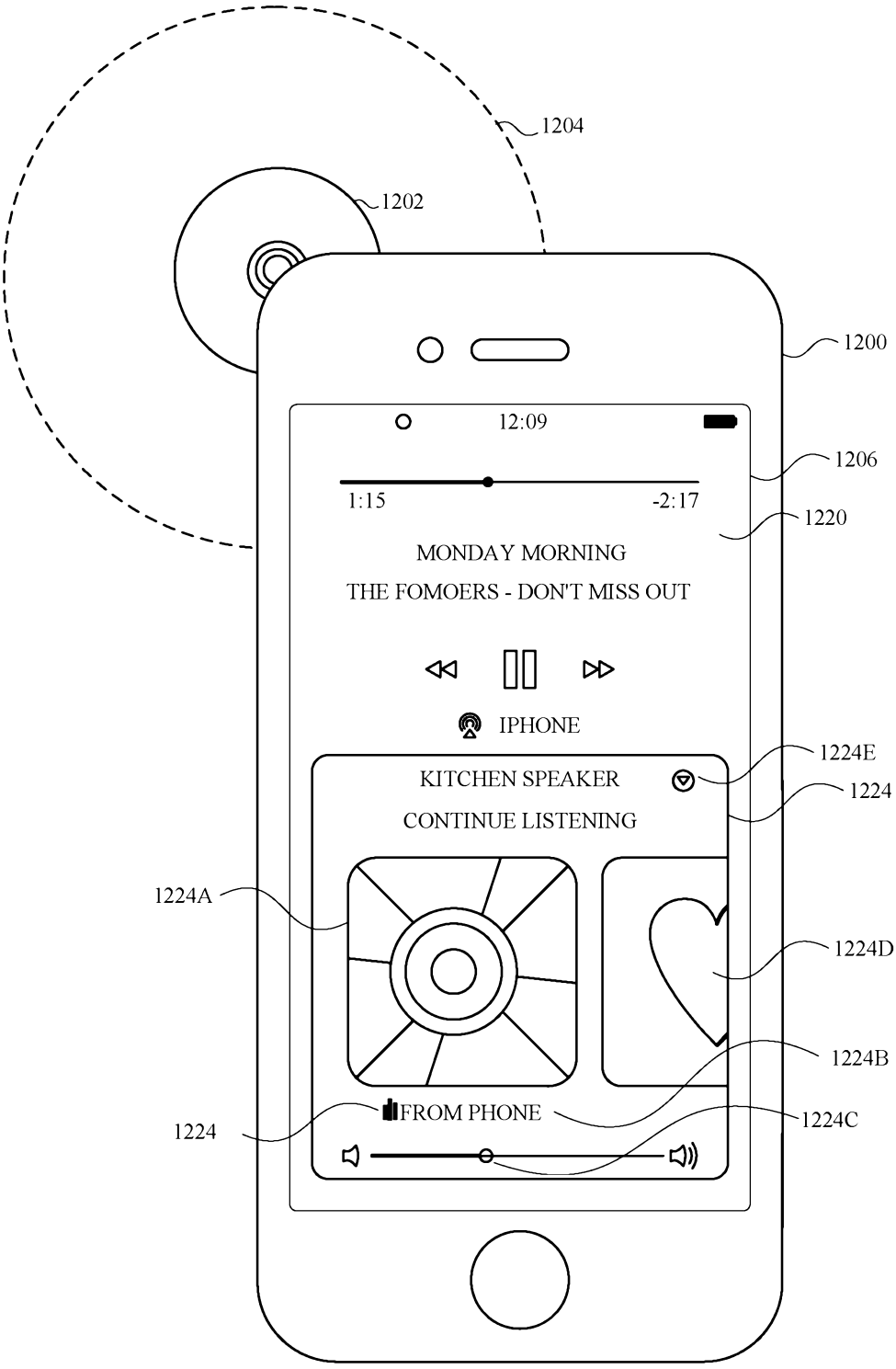

1204
1202
1200
12:09
1206
1:15
-2:17
1220
MONDAY MORNING
THE FOMOERS - DON'T MISS OUT
IPHONE
1224E
KITCHEN SPEAKER
1224
CONTINUE LISTENING
1224A
1224D
1224B
FROM PHONE
1224C
1224

FIG. 12I

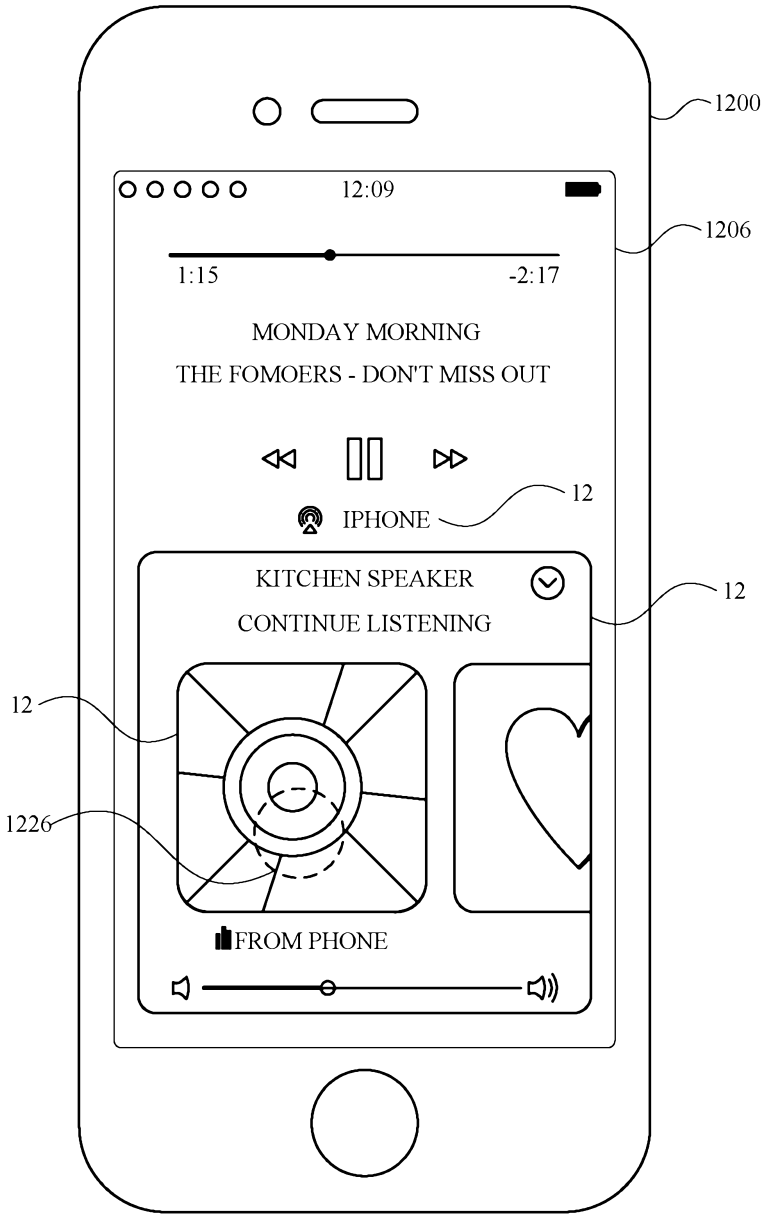

1200
12:09
1206
1:15
-2:17
MONDAY MORNING
THE FOMOERS - DON'T MISS OUT
12
IPHONE
KITCHEN SPEAKER
CONTINUE LISTENING
12
12
1226
FROM PHONE

FIG. 12J

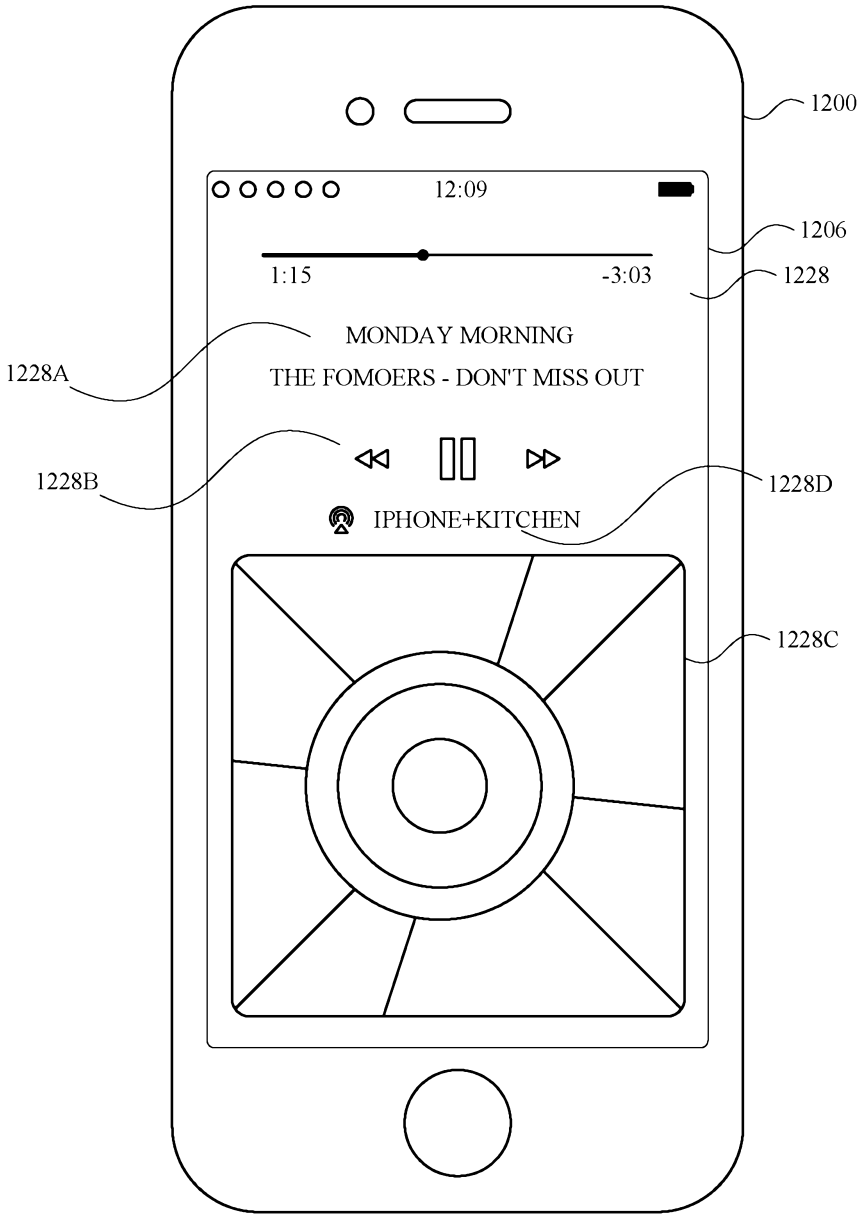

1200
12:09
1206
1:15
-3:03
1228
MONDAY MORNING
THE FOMOERS - DON'T MISS OUT
1228A
1228B
1228D
IPHONE+KITCHEN
1228C

FIG. 12K

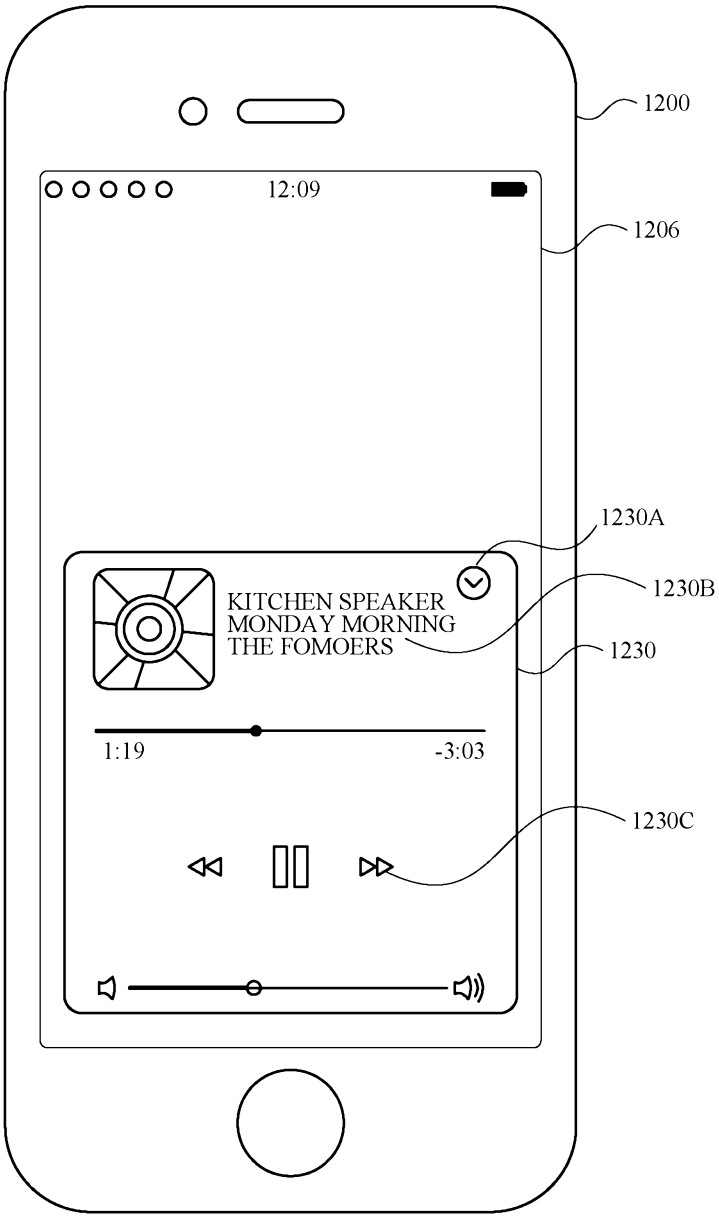

1200
12:09
1206
1230A
1230B
KITCHEN SPEAKER
MONDAY MORNING
THE FOMOERS
1230
1:19
-3:03
1230C

FIG. 12L

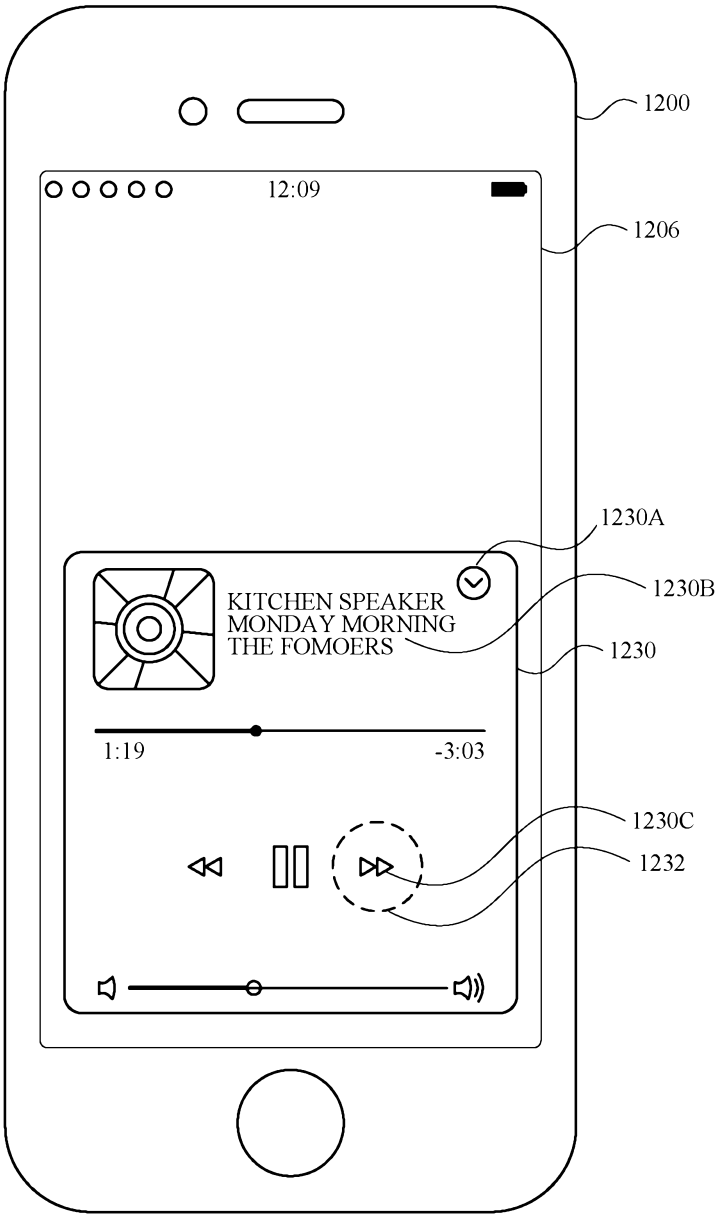

1200
12:09
1206
1230A
1230B
KITCHEN SPEAKER
MONDAY MORNING
THE FOMOERS
1230
1:19
-3:03
1230C
1232

FIG. 12M

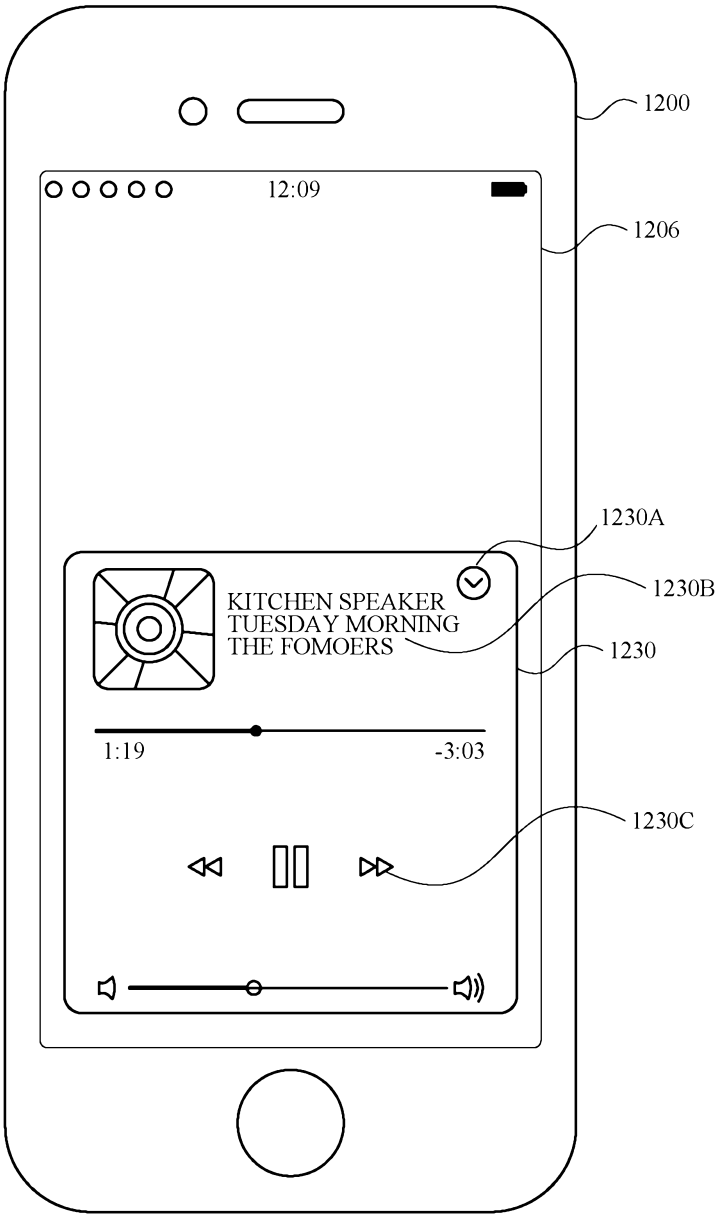

1200
12:09
1206
1230A
1230B
KITCHEN SPEAKER
TUESDAY MORNING
THE FOMOERS
1230
1:19
-3:03
1230C

FIG. 12N

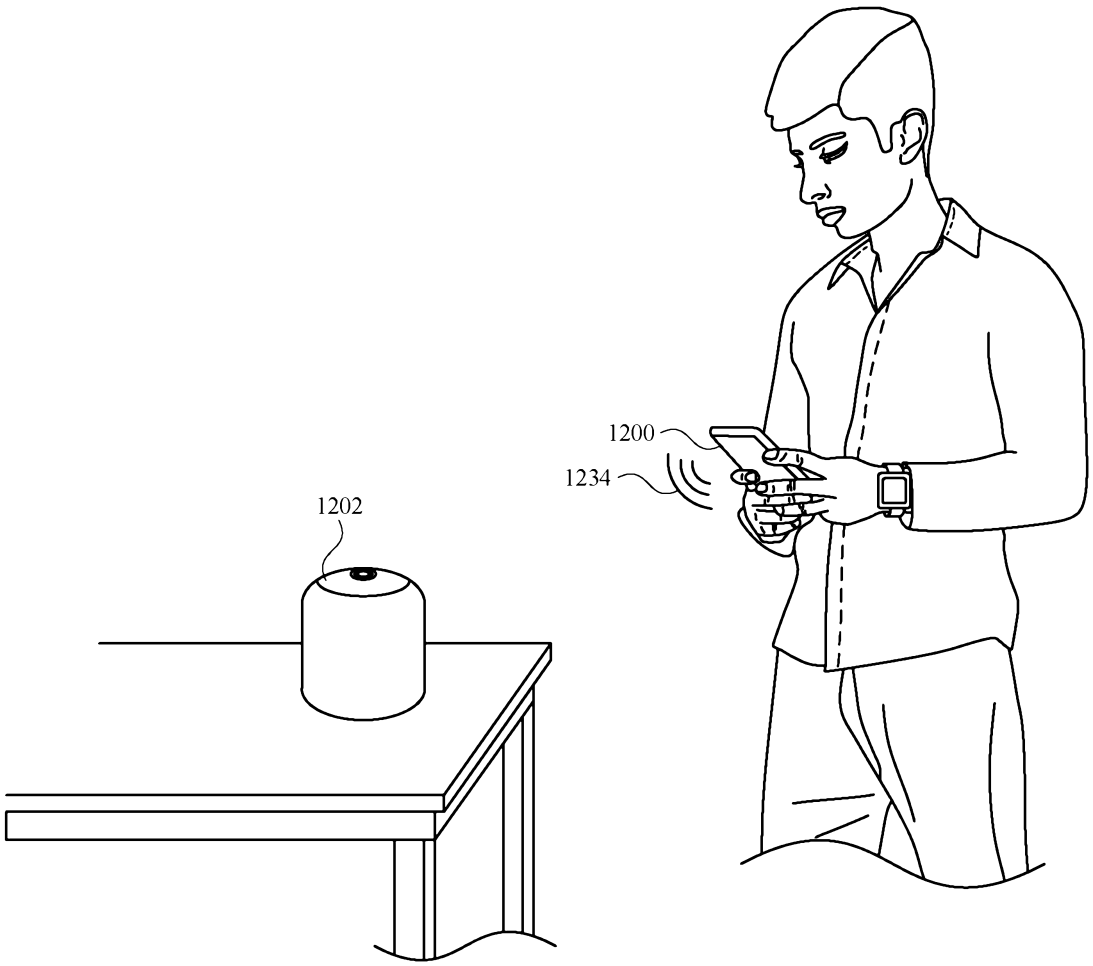


FIG. 12O

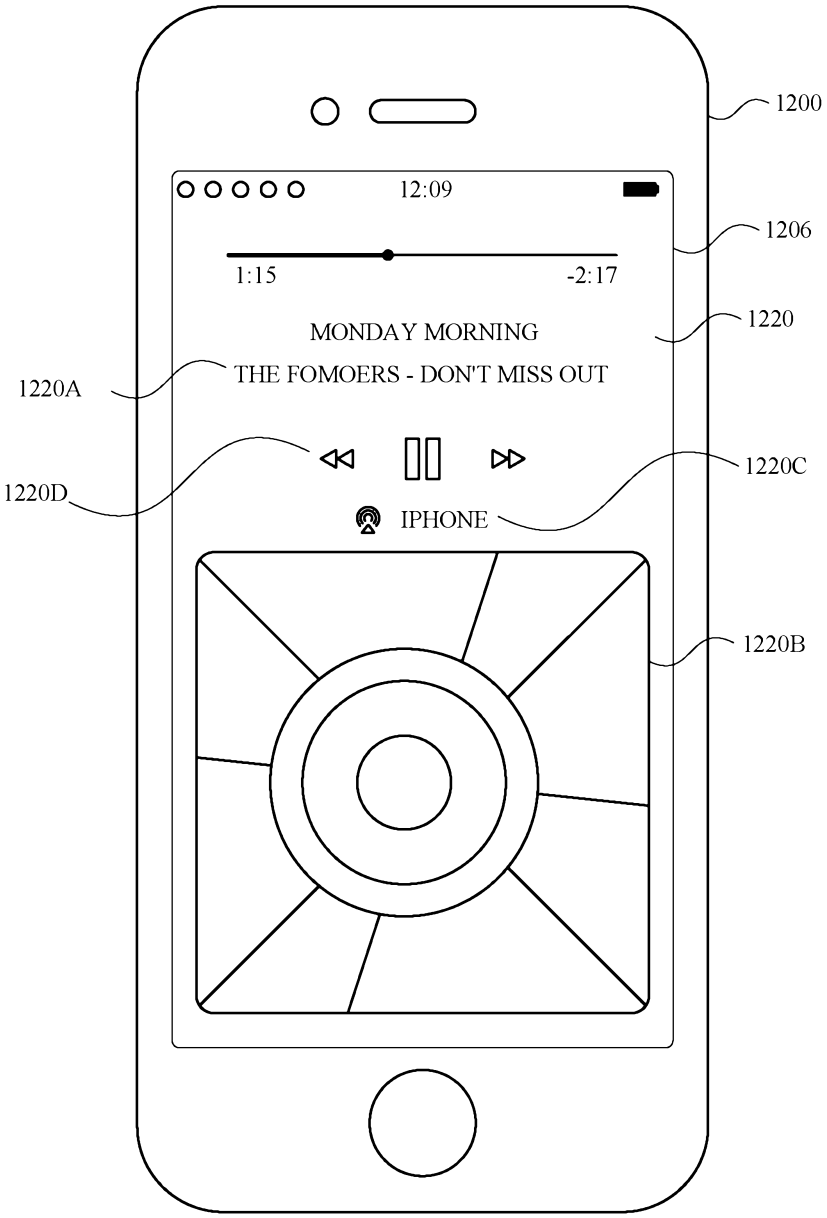

1200
12:09
1206
1:15
-2:17
MONDAY MORNING
1220
THE FOMOERS - DON'T MISS OUT
1220A
1220D
1220C
IPHONE
1220B

FIG. 12P

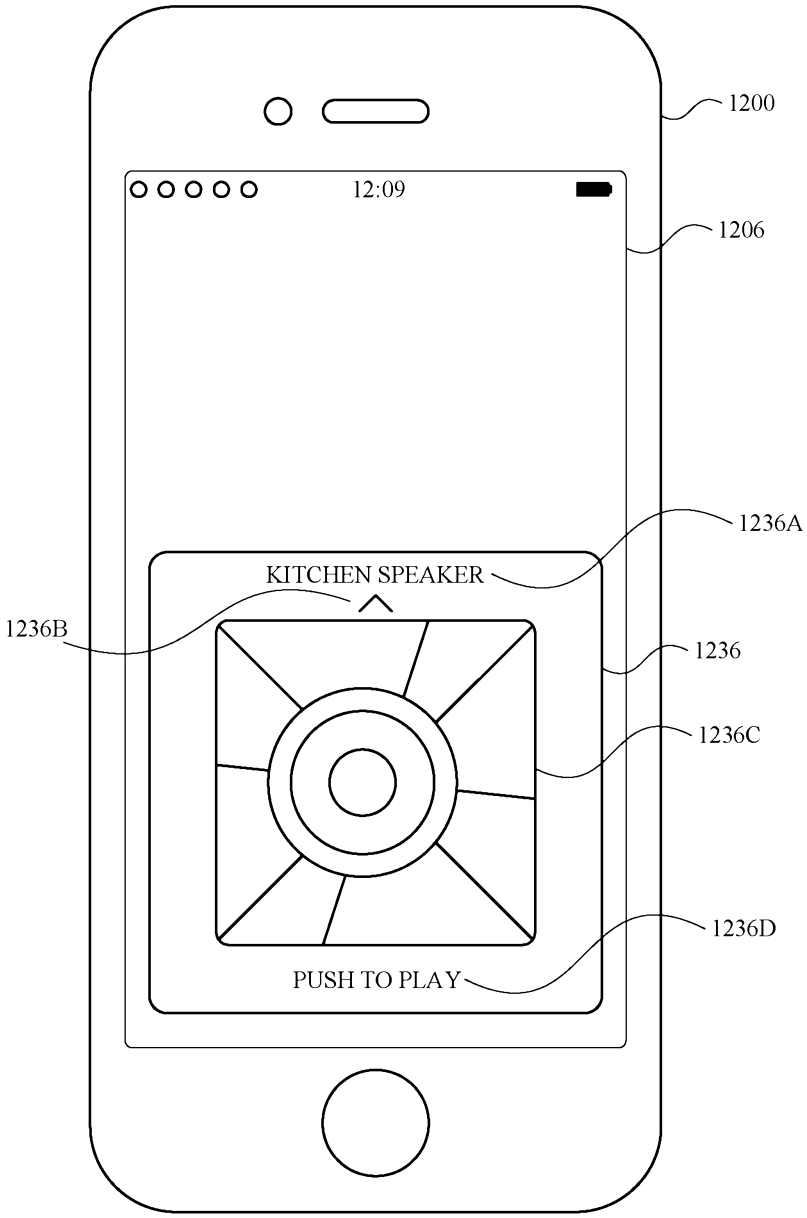

1200
12:09
1206
1236A
KITCHEN SPEAKER
1236B
1236
1236C
1236D
PUSH TO PLAY

FIG. 12Q

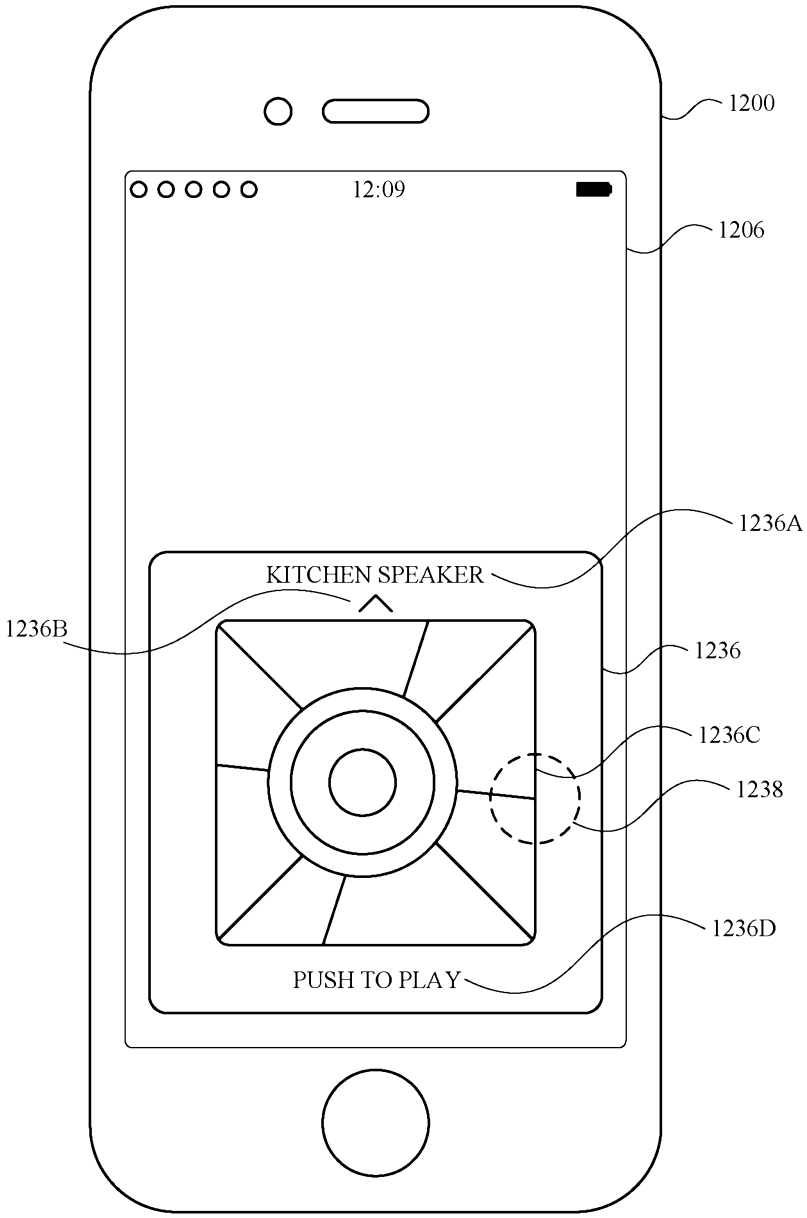

1200
12:09
1206
1236A
KITCHEN SPEAKER
1236B
1236
1236C
1238
1236D
PUSH TO PLAY

FIG. 12R

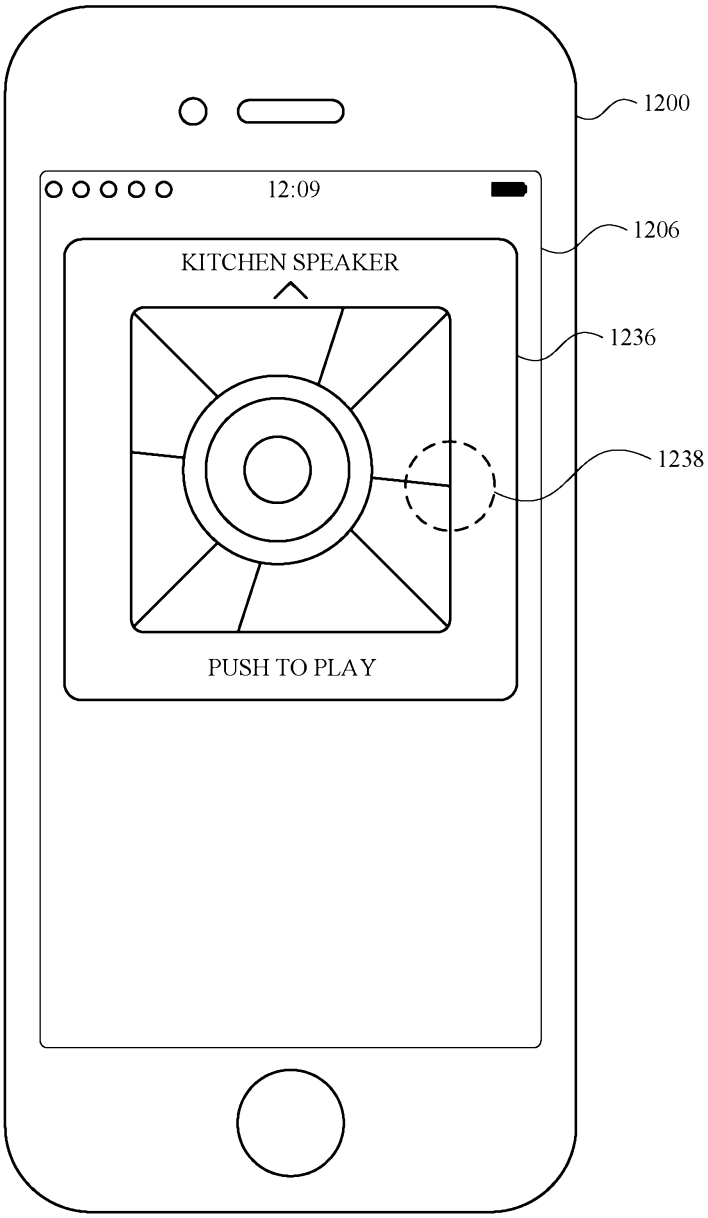


FIG. 12S

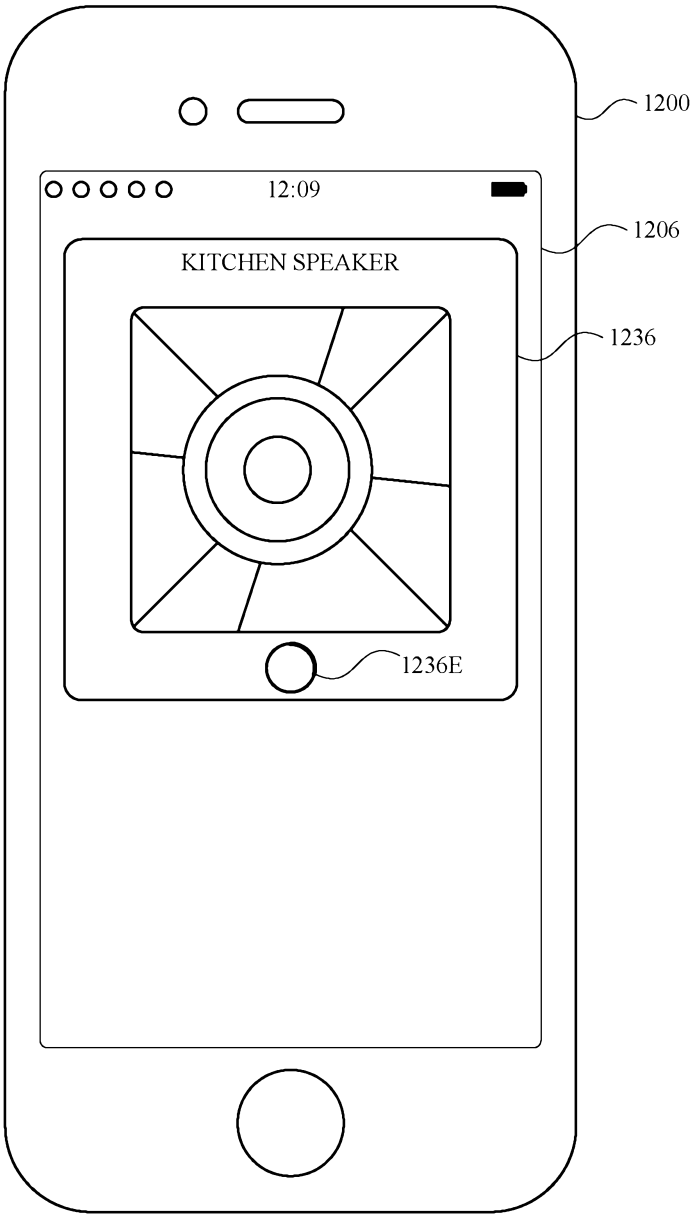


FIG. 12T

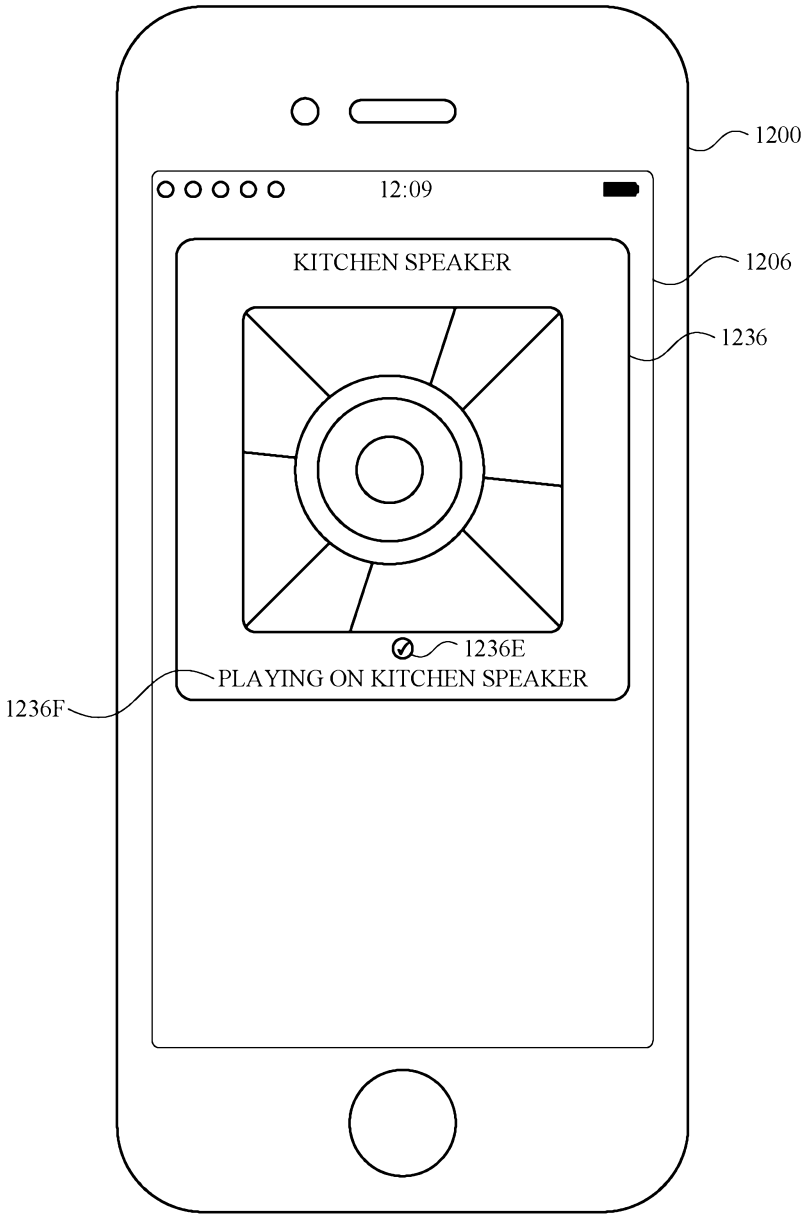

1200
12:09
KITCHEN SPEAKER
1206
1236
1236E
PLAYING ON KITCHEN SPEAKER
1236F

FIG. 12U

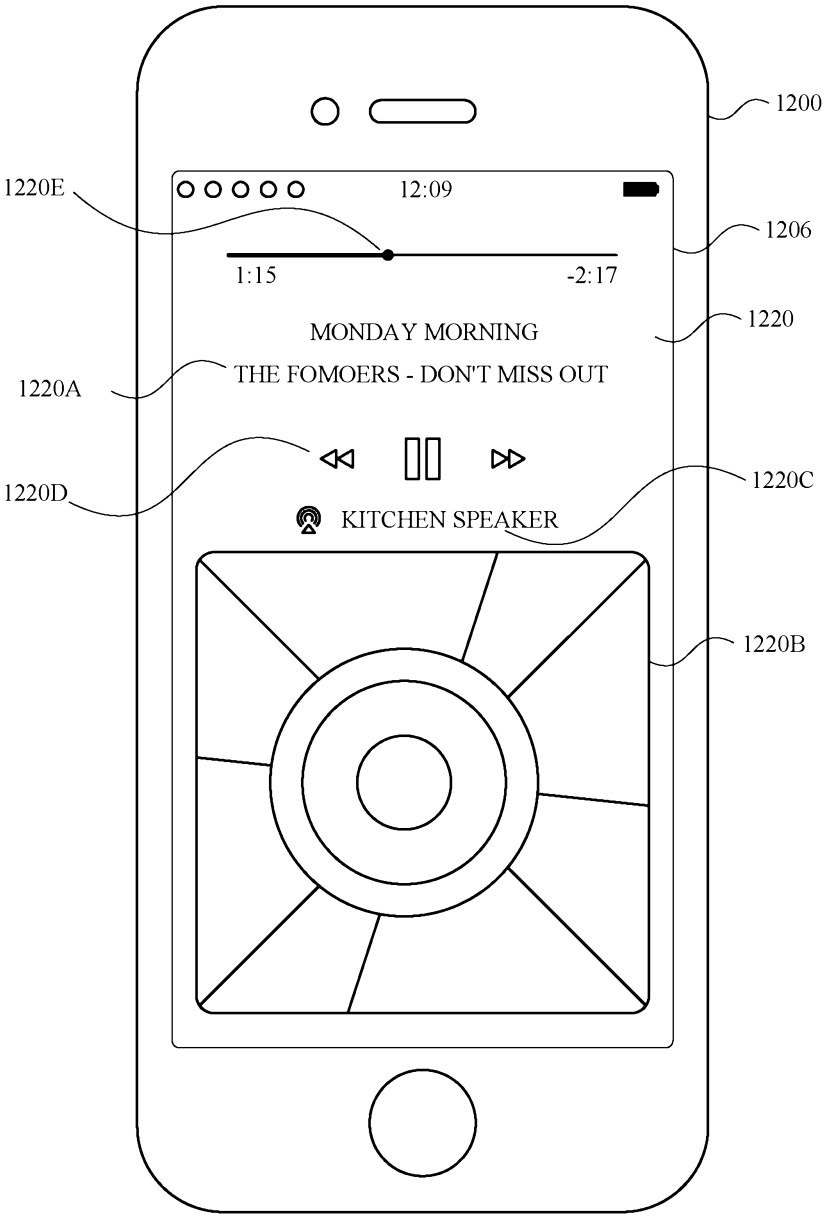

1200
1220E
12:09
1206
1:15
-2:17
1220
MONDAY MORNING
1220A
THE FOMOERS - DON'T MISS OUT
1220D
1220C
KITCHEN SPEAKER
1220B

FIG. 12V

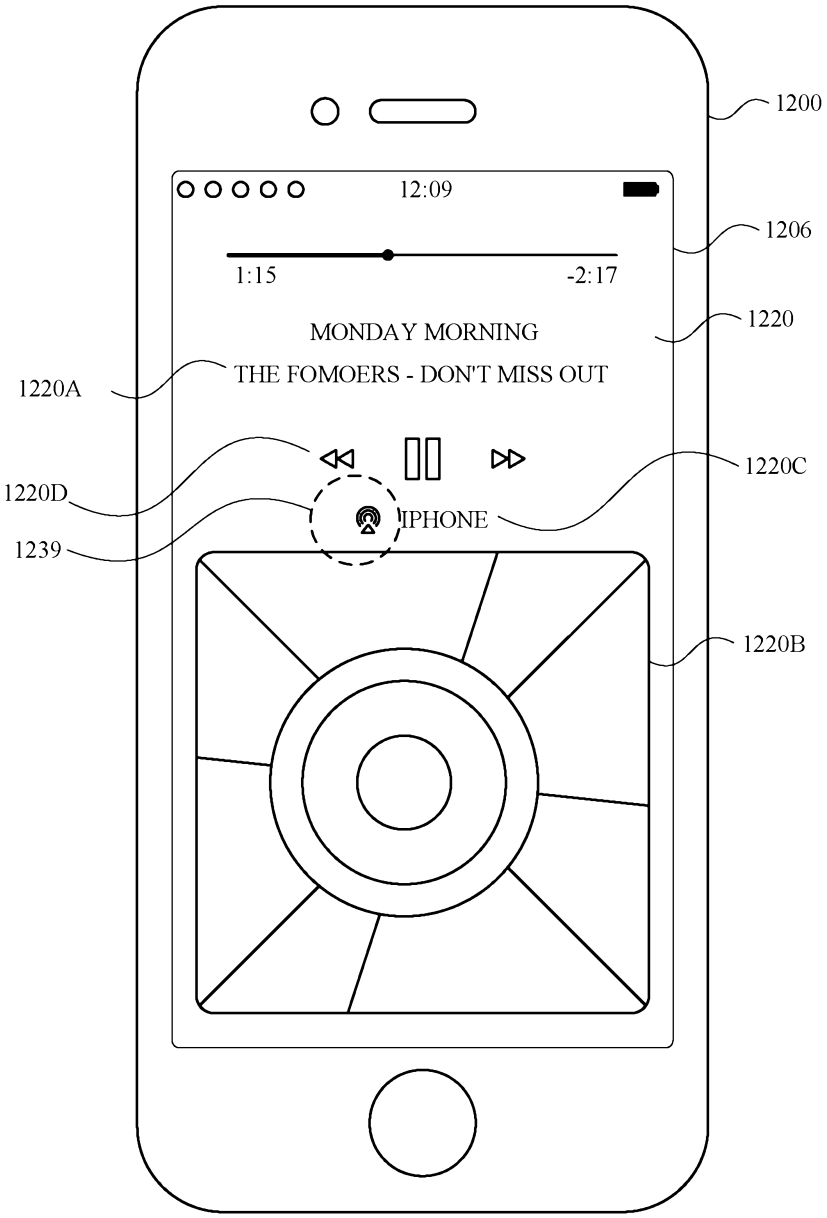

1200
12:09
1206
1:15
-2:17
MONDAY MORNING
1220
THE FOMOERS - DON'T MISS OUT
1220A
1220C
1220D
IPHONE
1239
1220B

FIG. 12W

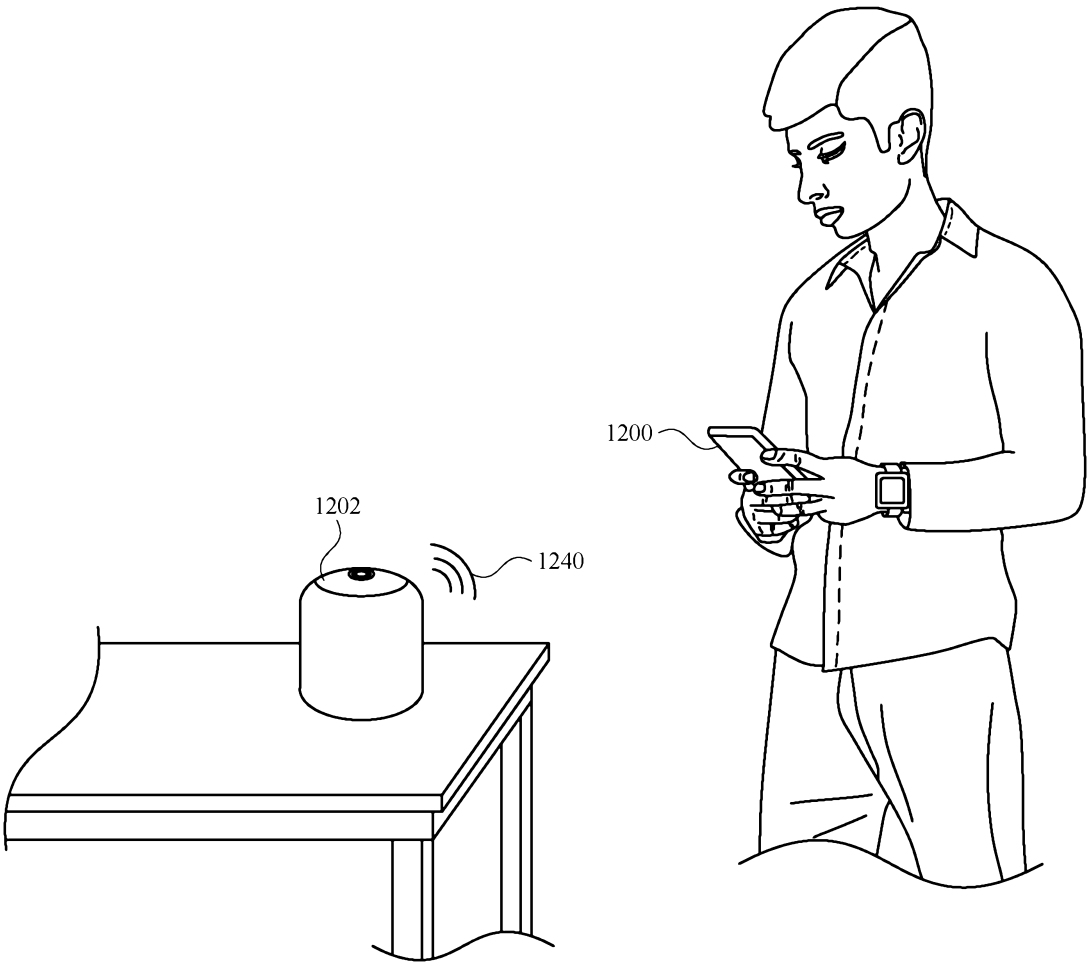


FIG. 12X

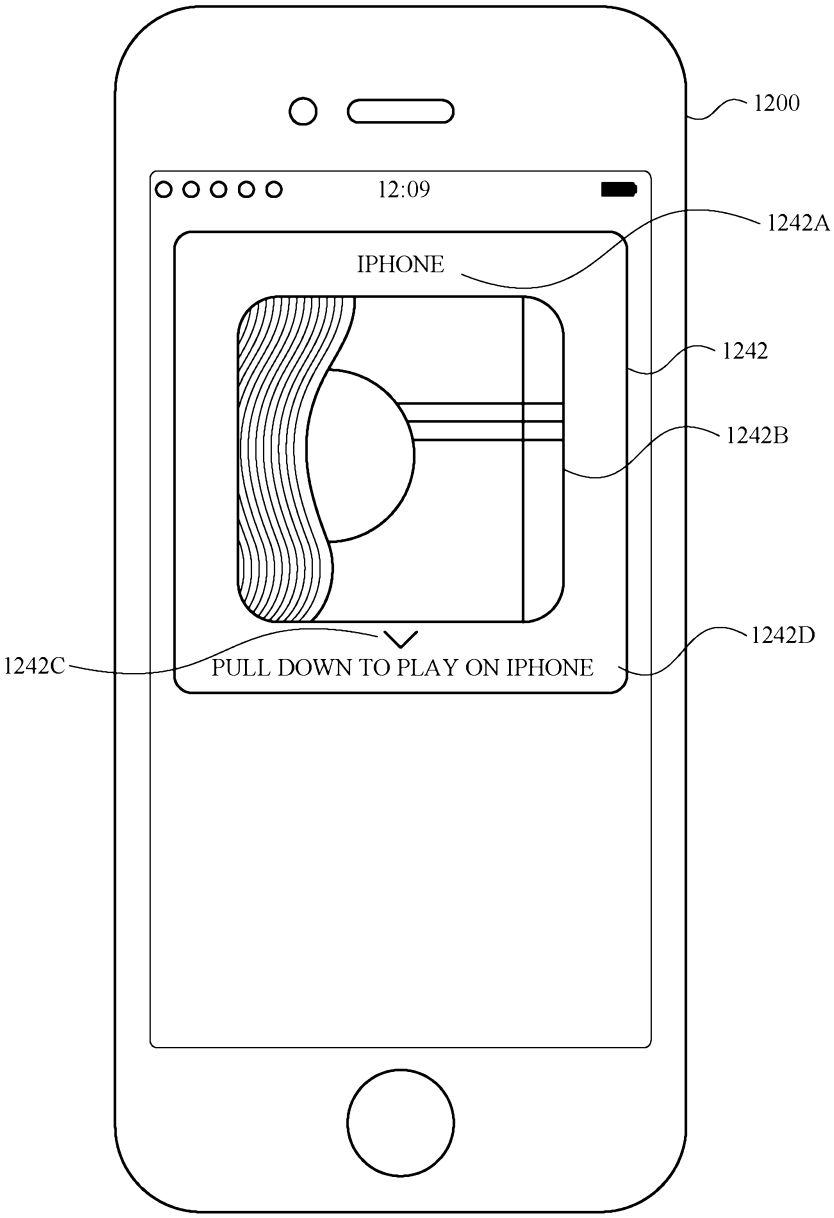

1200
12:09
1242A
IPHONE
1242
1242B
1242C
PULL DOWN TO PLAY ON IPHONE
1242D

FIG. 12Y

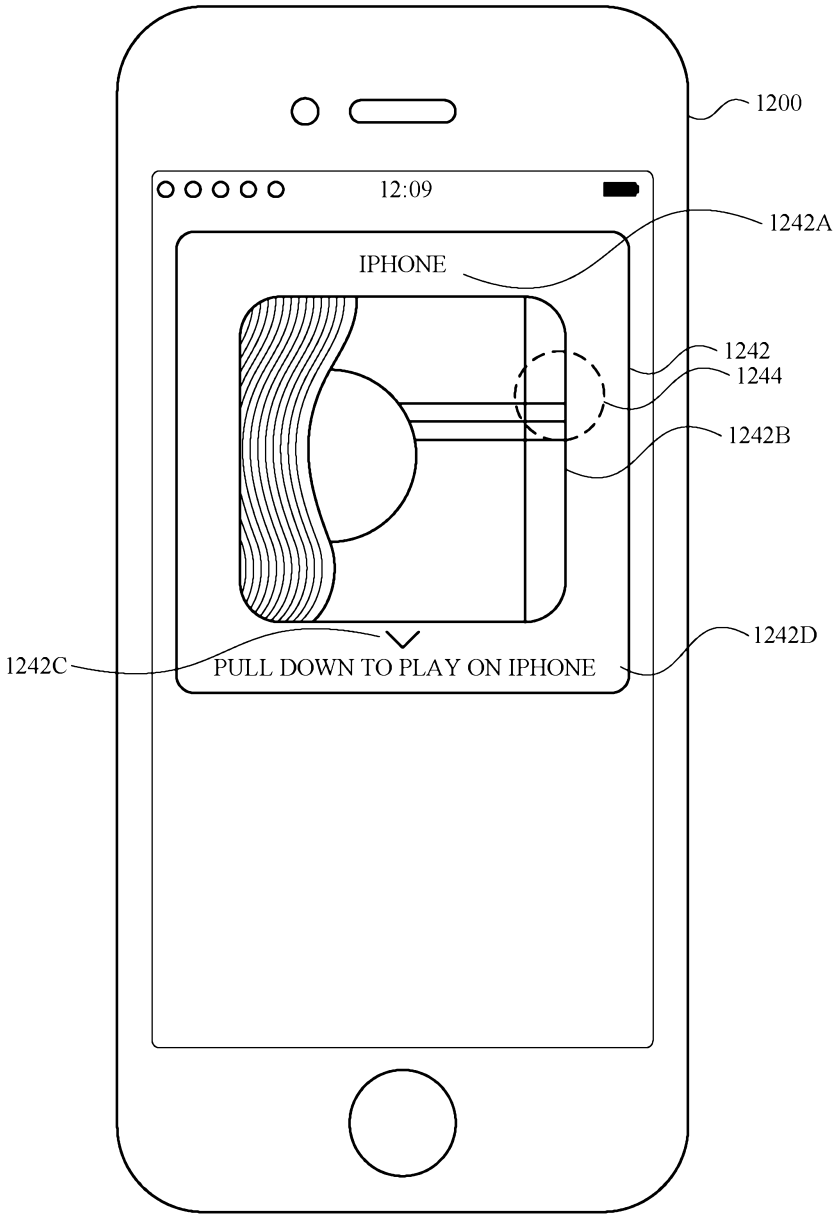

1200
12:09
1242A
IPHONE
1242
1244
1242B
1242C
PULL DOWN TO PLAY ON IPHONE
1242D

FIG. 12Z

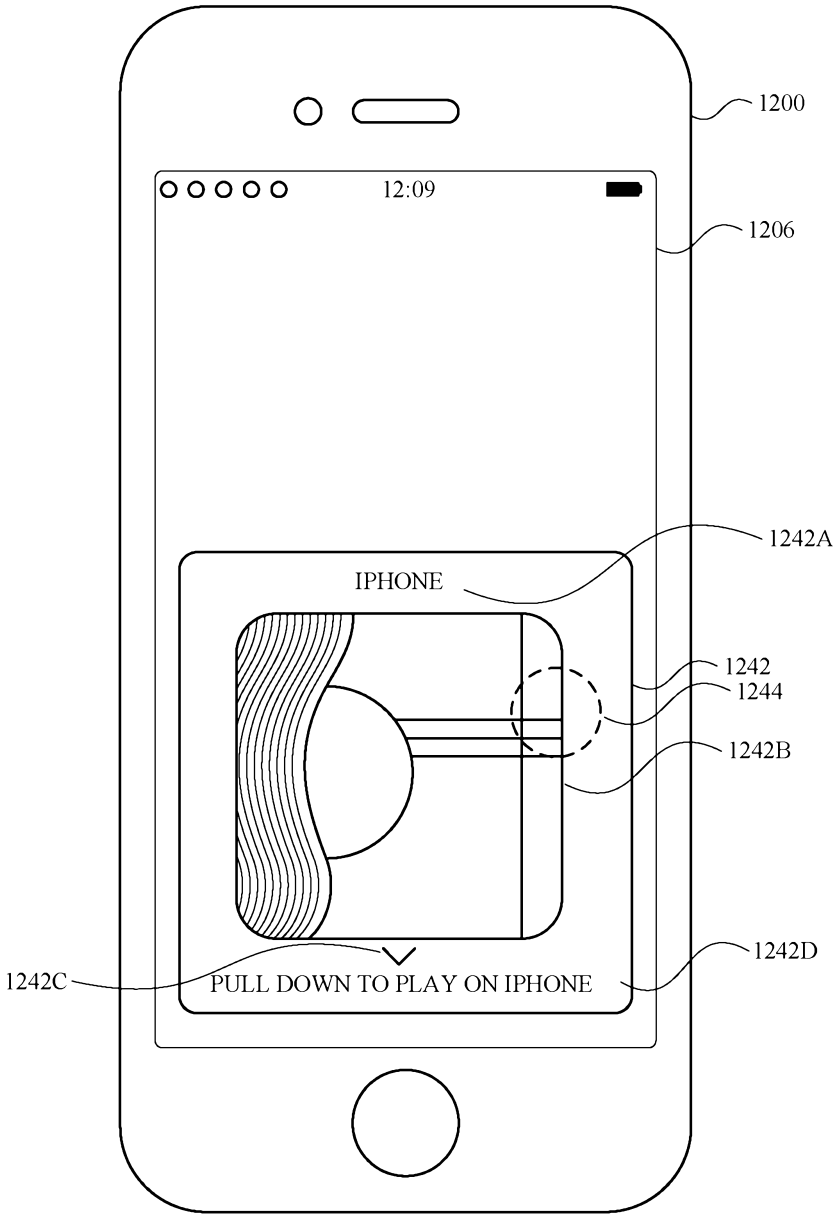

1200
12:09
1206
1242A
IPHONE
1242
1244
1242B
1242C
PULL DOWN TO PLAY ON IPHONE
1242D

FIG. 12AA

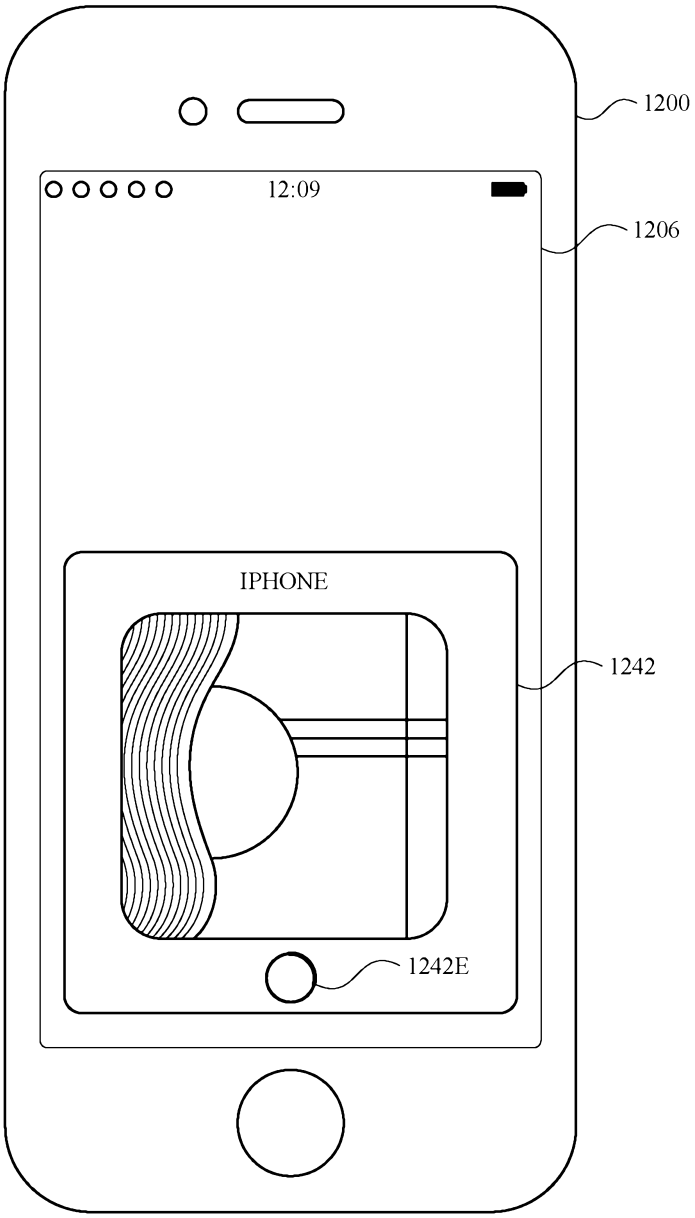


FIG. 12AB

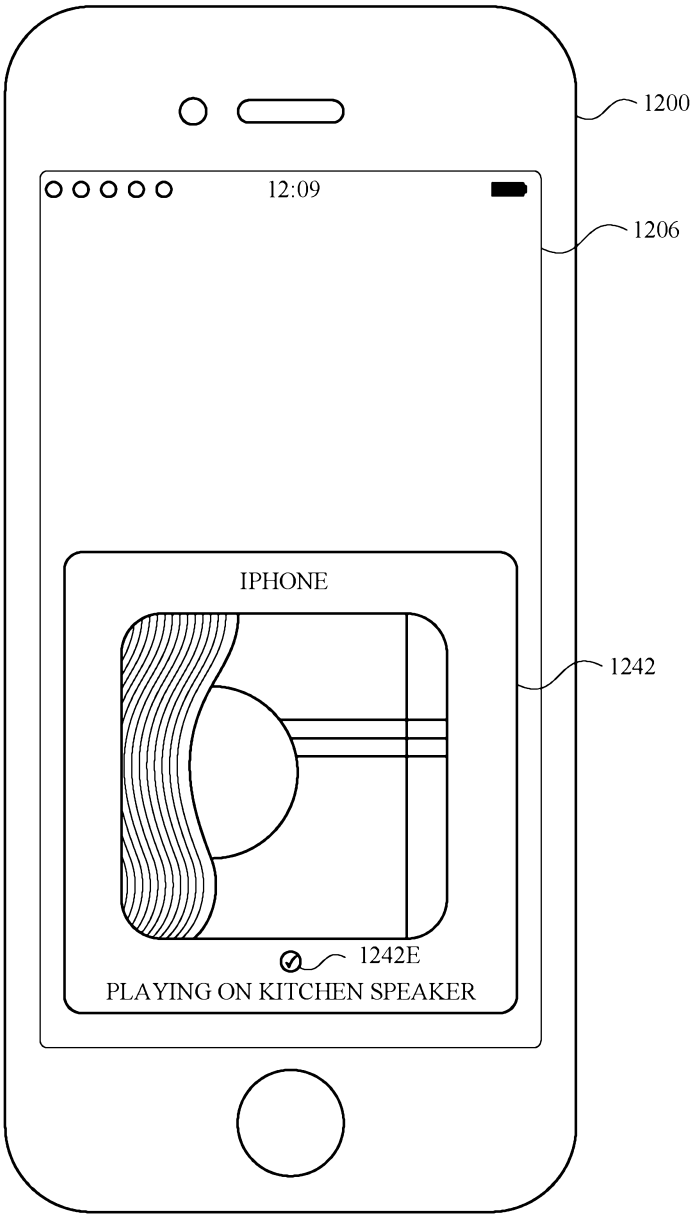

1200
12:09
1206
IPHONE
1242
1242E
PLAYING ON KITCHEN SPEAKER

FIG. 12AC

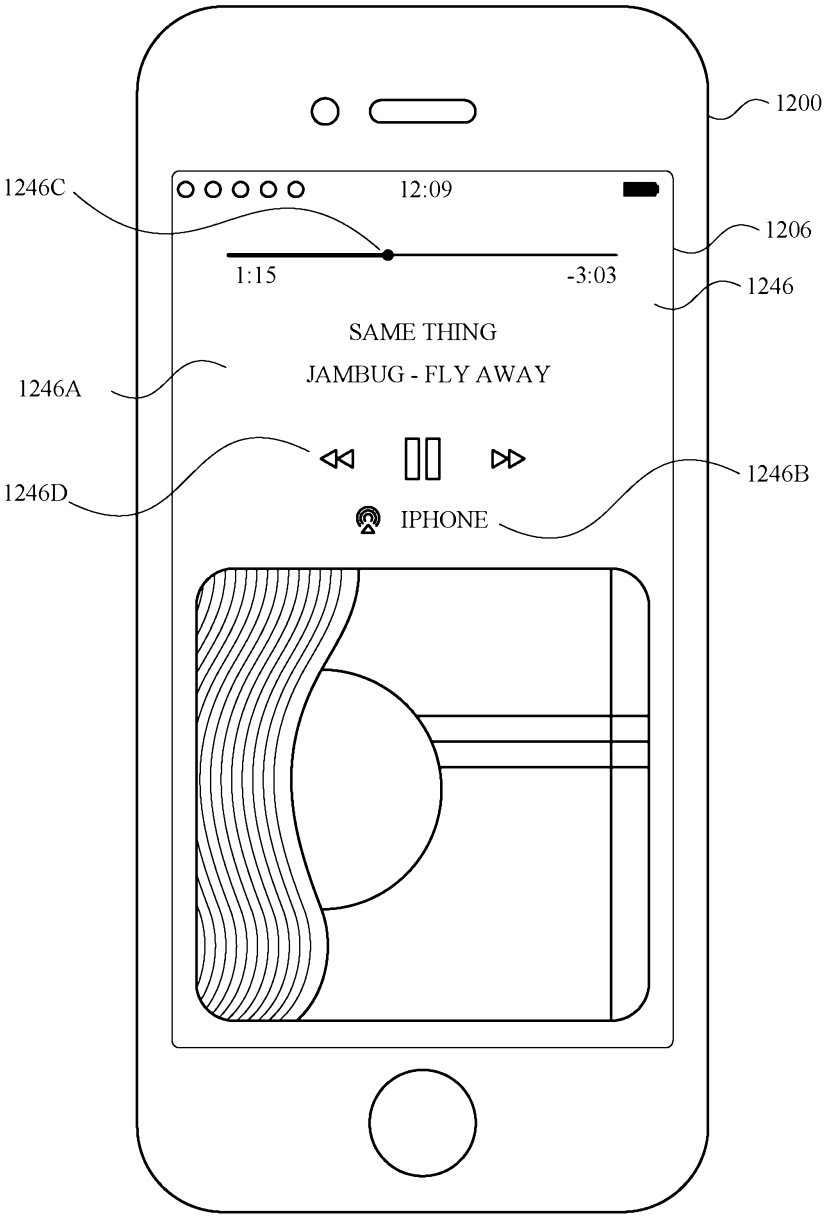

1200
1246C
12:09
1206
1:15
-3:03
1246
SAME THING
JAMBUG - FLY AWAY
1246A
1246D
1246B
IPHONE

FIG. 12AD

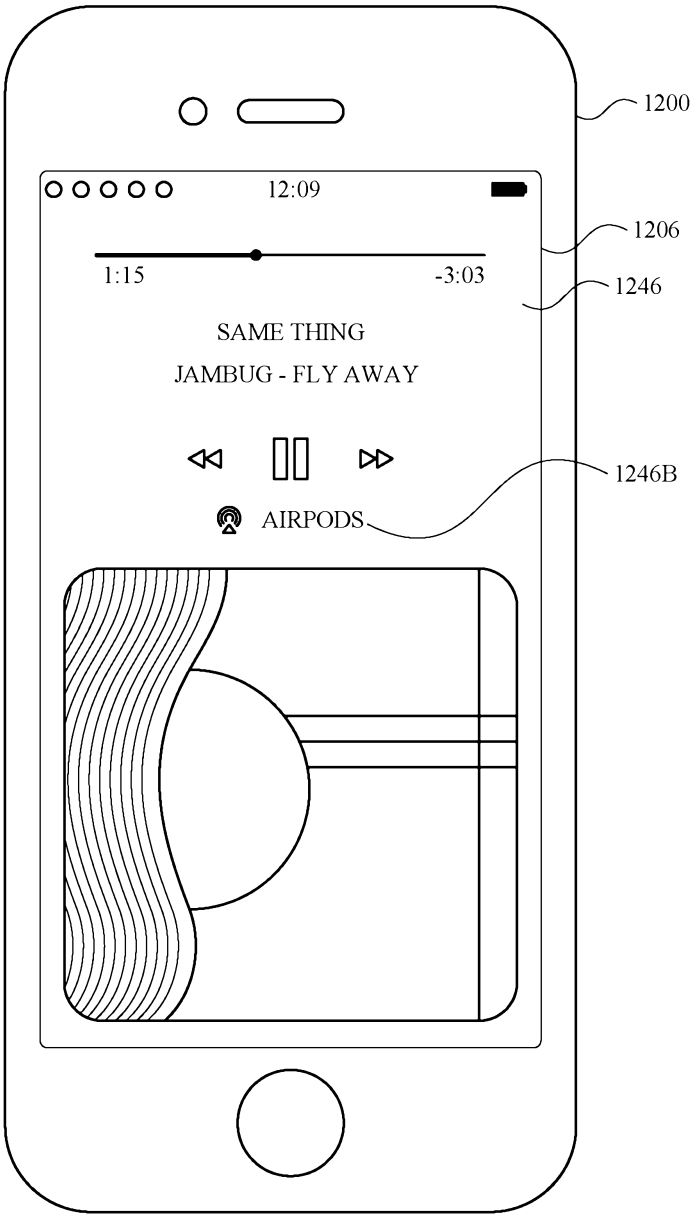

1200
12:09
1206
1:15
-3:03
1246
SAME THING
JAMBUG - FLY AWAY
1246B
AIRPODS

FIG. 12AE

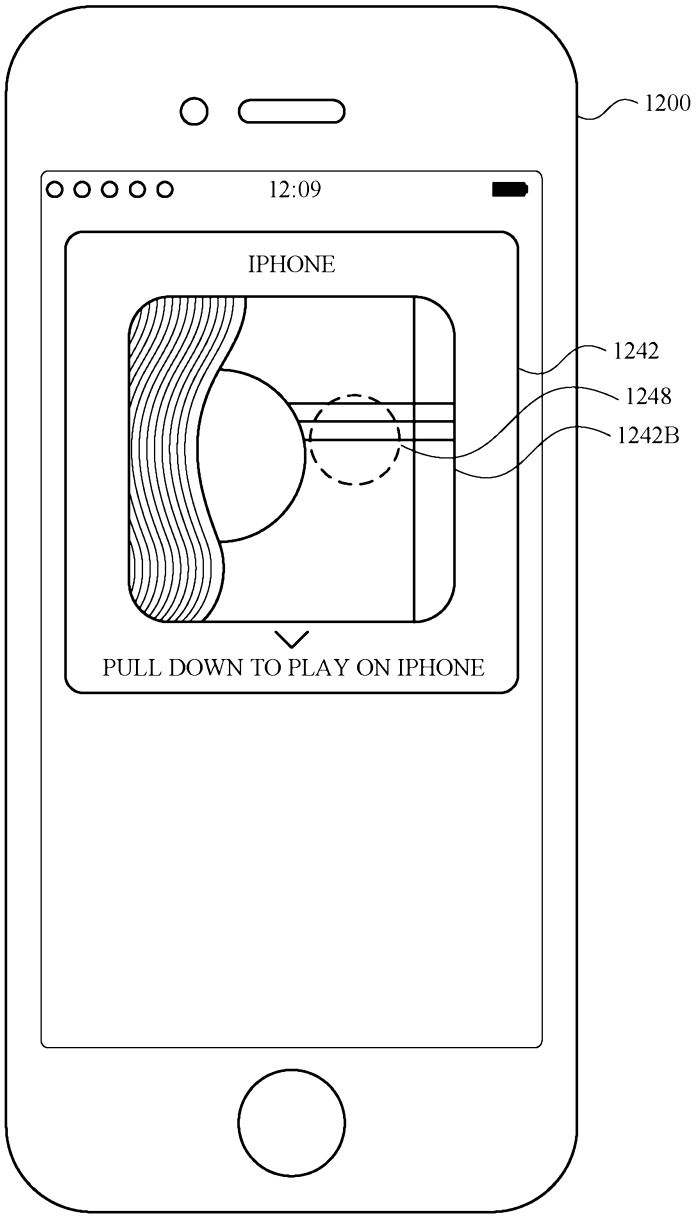


FIG. 12AF

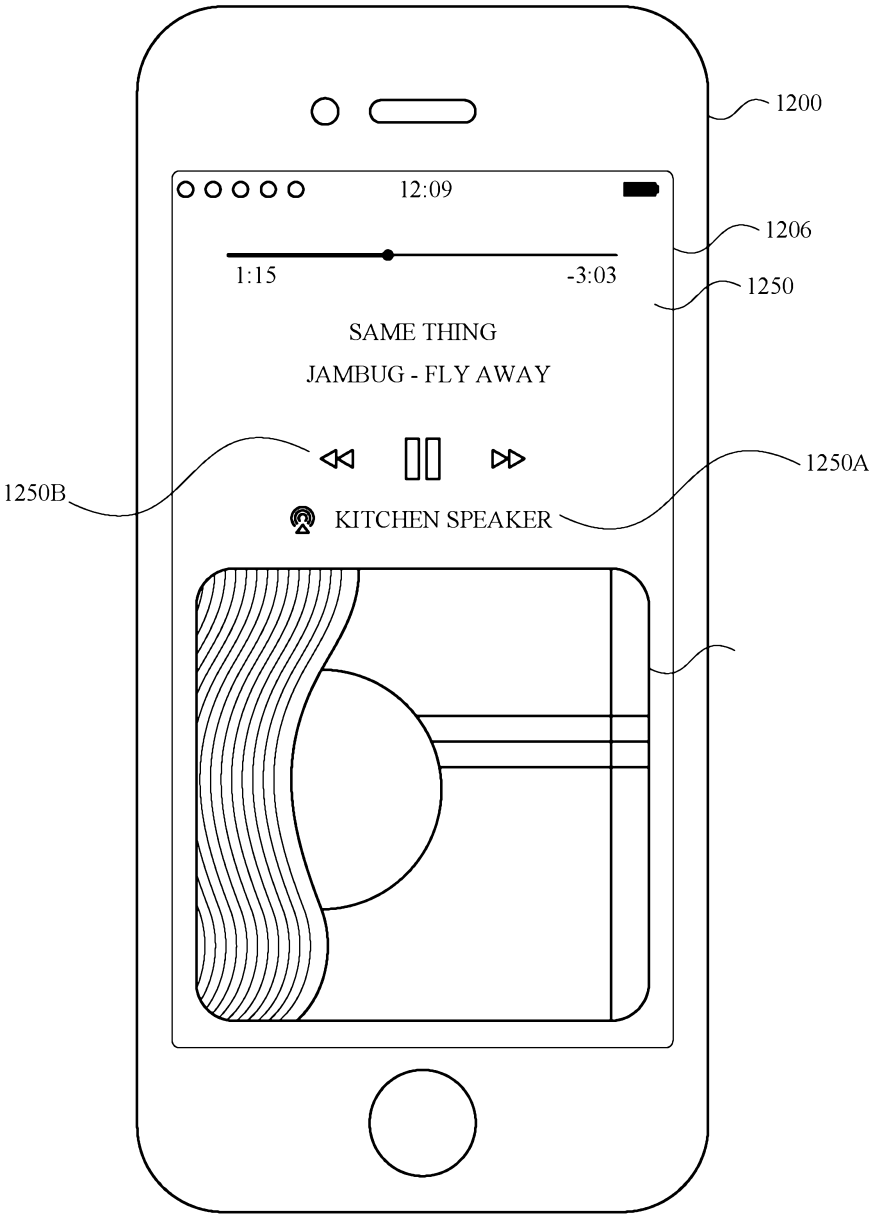

1200
12:09
1206
1:15
-3:03
1250
SAME THING
JAMBUG - FLY AWAY
1250B
1250A
KITCHEN SPEAKER

FIG. 12AG

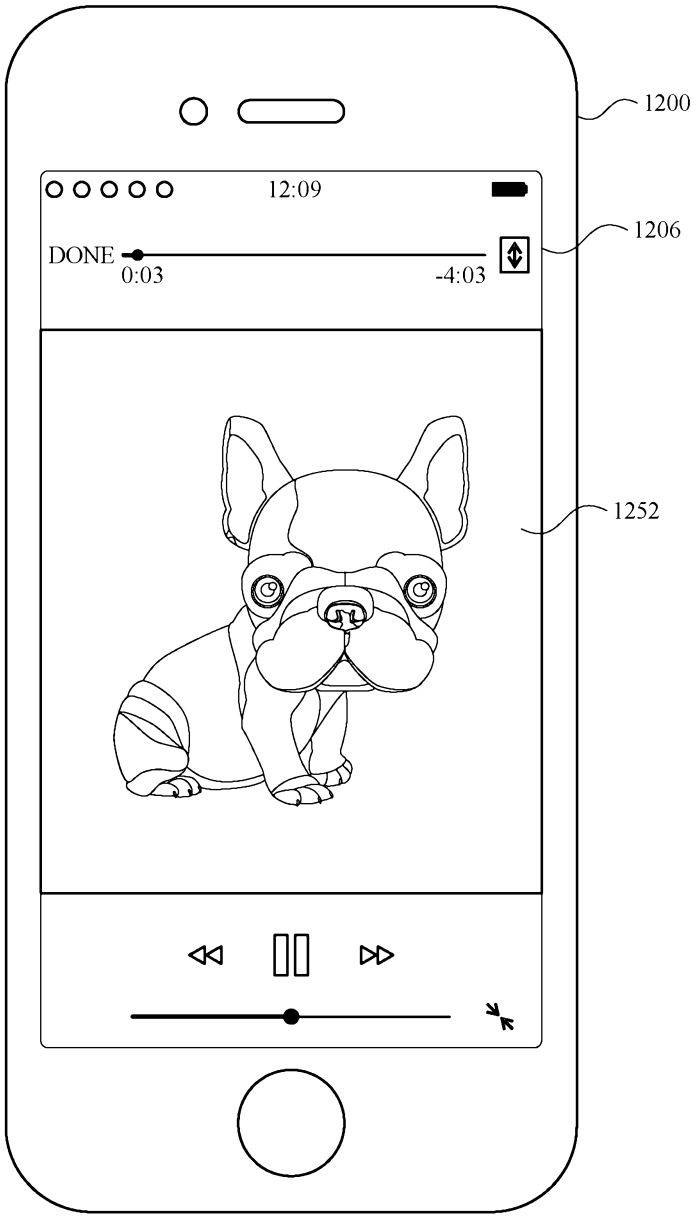


FIG. 12AH

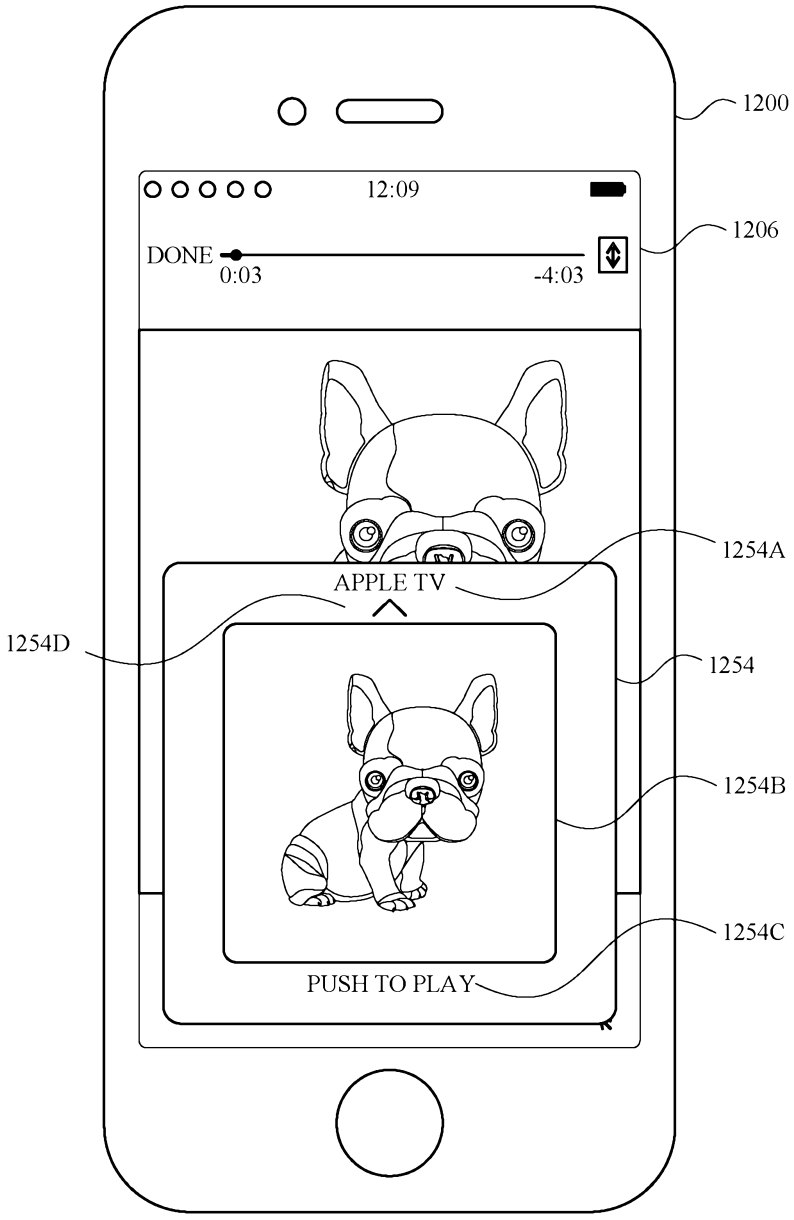

1200
12:09
1206
DONE
0:03
-4:03
1254A
APPLE TV
1254D
1254
1254B
1254C
PUSH TO PLAY

FIG. 12AI

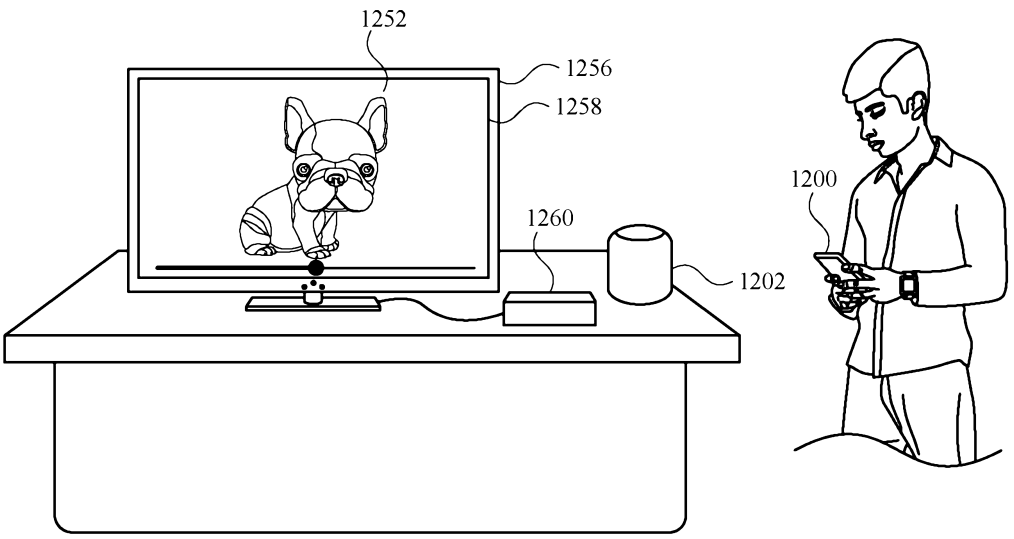


FIG. 12AJ

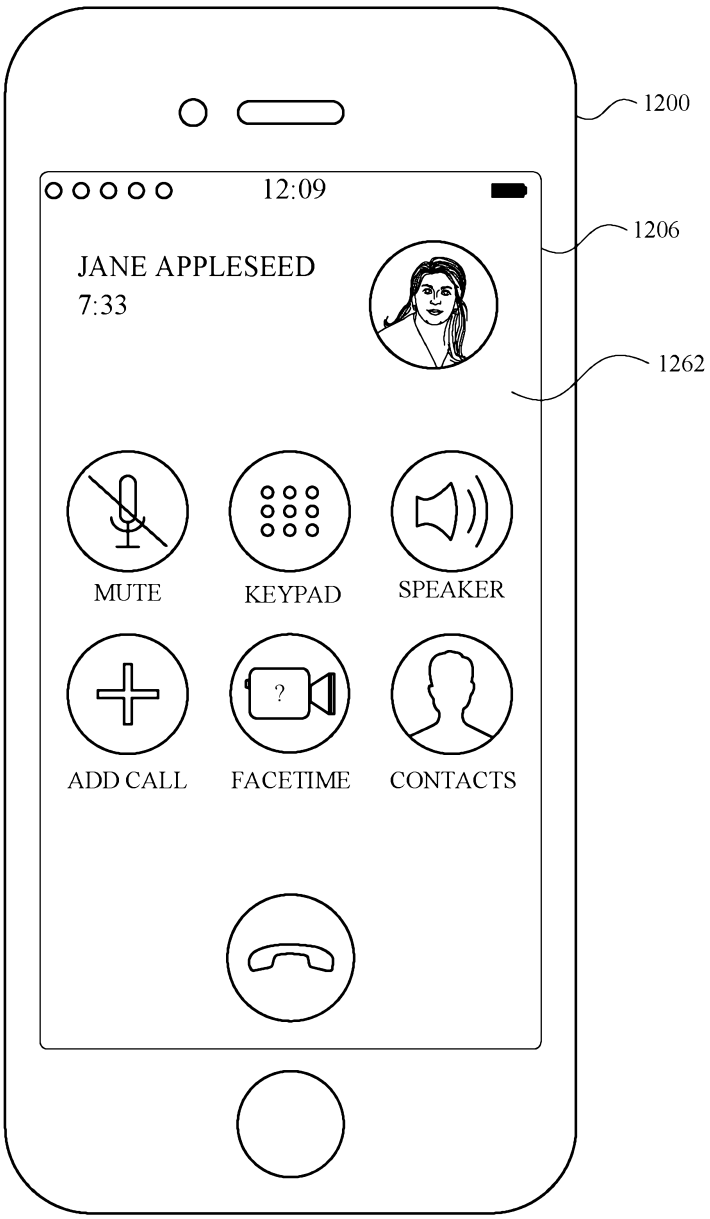

1200
12:09
1206
JANE APPLESEED
7:33
1262
MUTE
KEYPAD
SPEAKER
ADD CALL
FACETIME
CONTACTS

FIG. 12AK

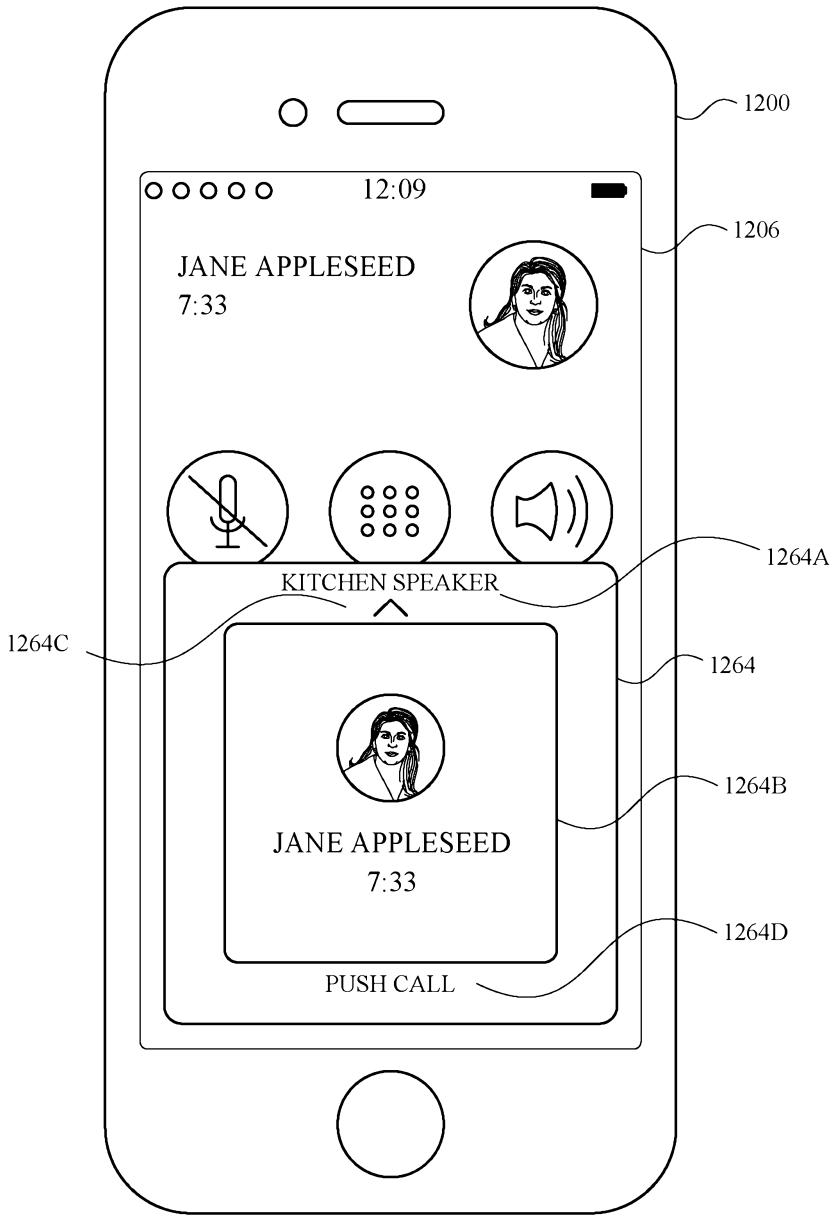

1200
12:09
1206
JANE APPLESEED
7:33
1264A
KITCHEN SPEAKER
1264C
1264
1264B
JANE APPLESEED
7:33
1264D
PUSH CALL

FIG. 12AL

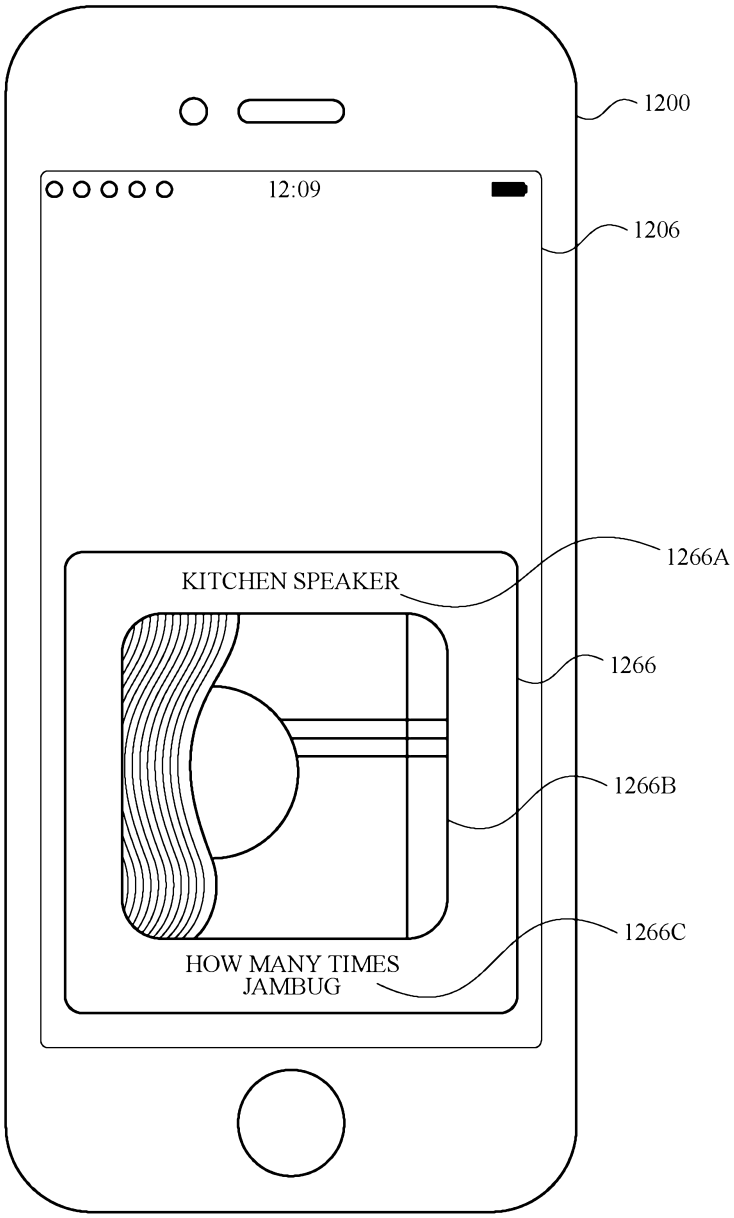

1200
12:09
1206
1266A
KITCHEN SPEAKER
1266
1266B
1266C
HOW MANY TIMES
JAMBUG

FIG. 12AM

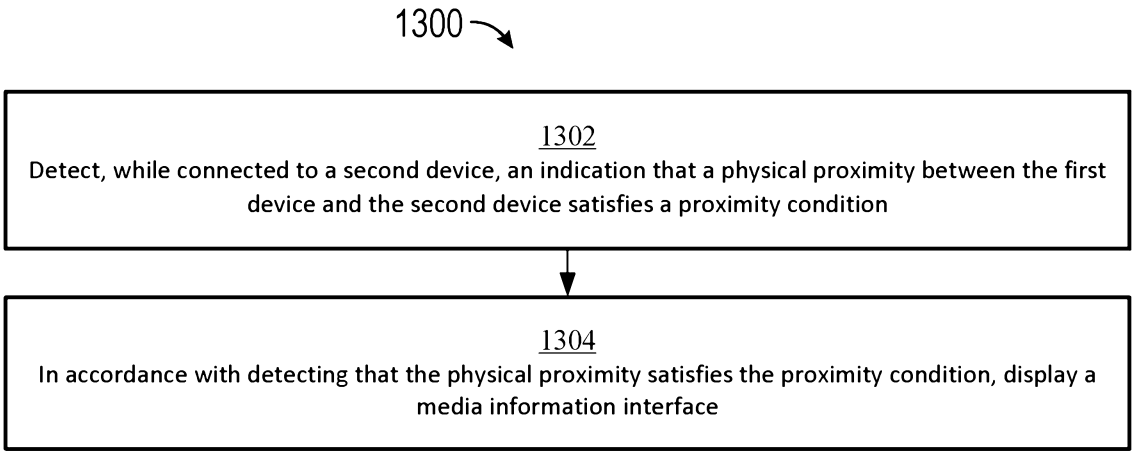

1300
1302
Detect, while connected to a second device, an indication that a physical proximity between the first device and the second device satisfies a proximity condition
1304
In accordance with detecting that the physical proximity satisfies the proximity condition, display a media information interface

FIG. 13

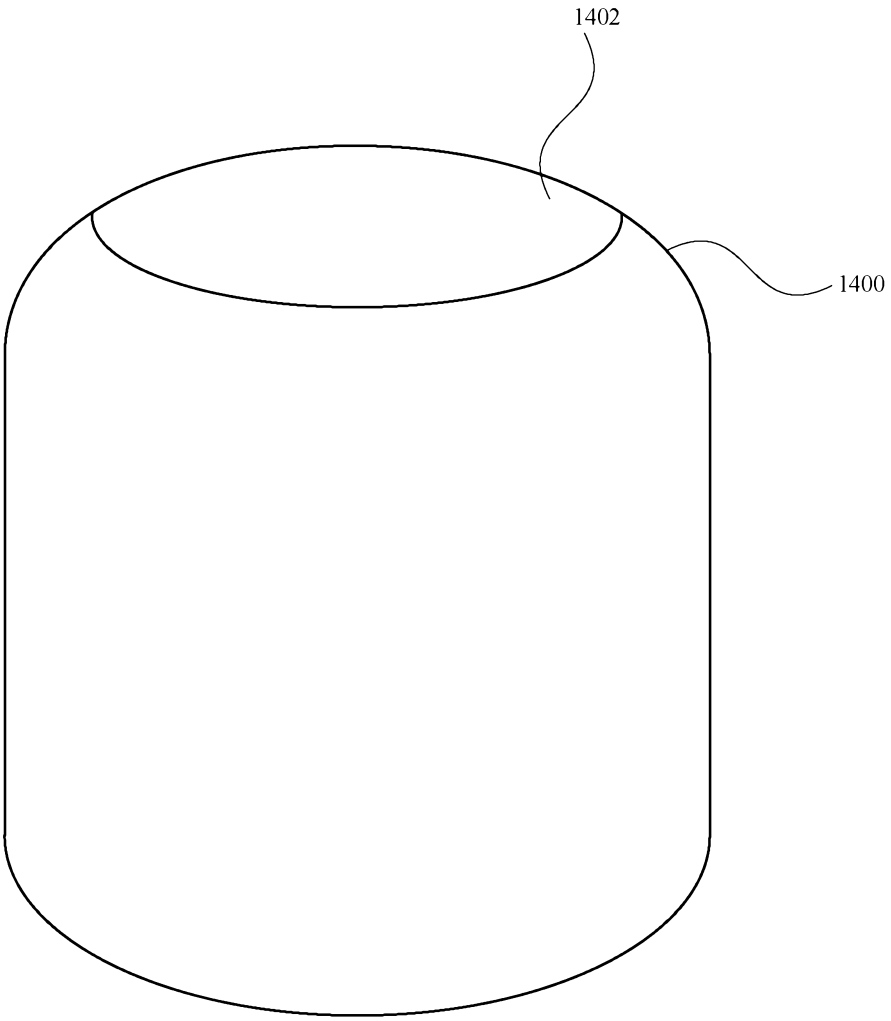


FIG. 14A

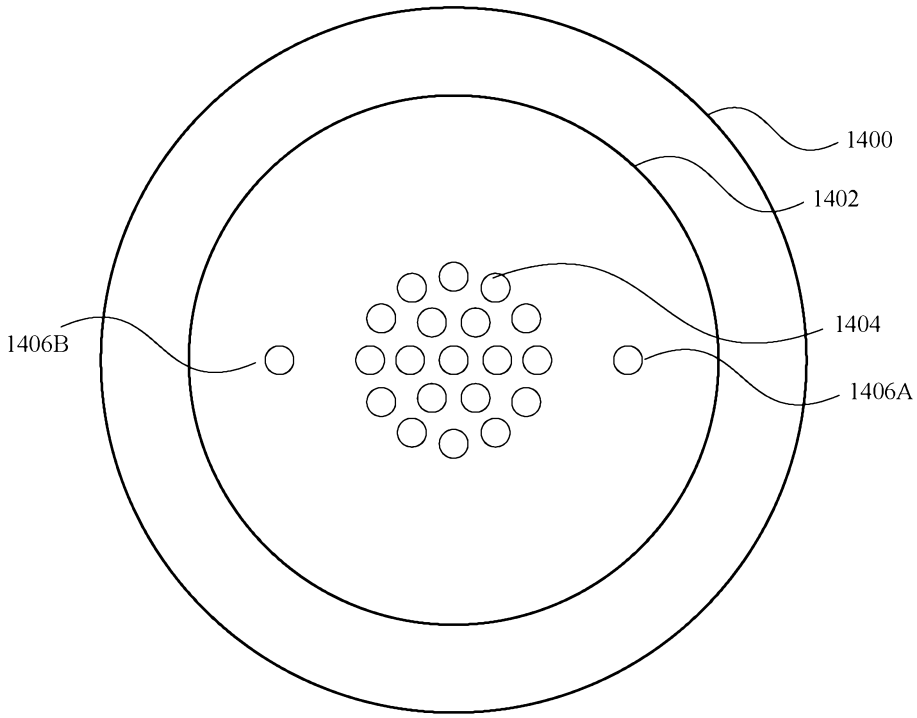


FIG. 14B

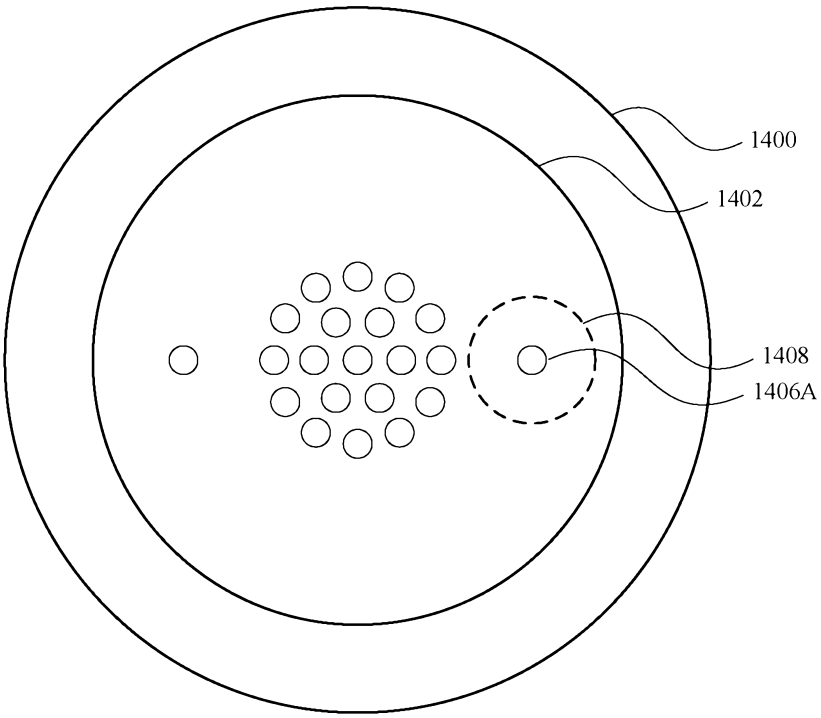


FIG. 14C

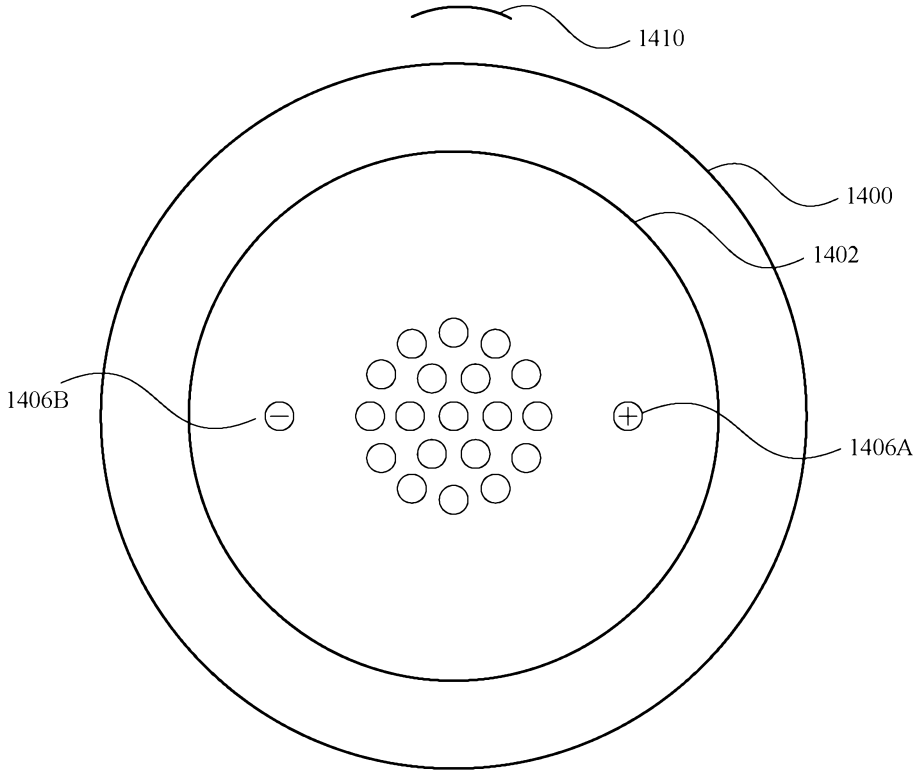


FIG. 14D

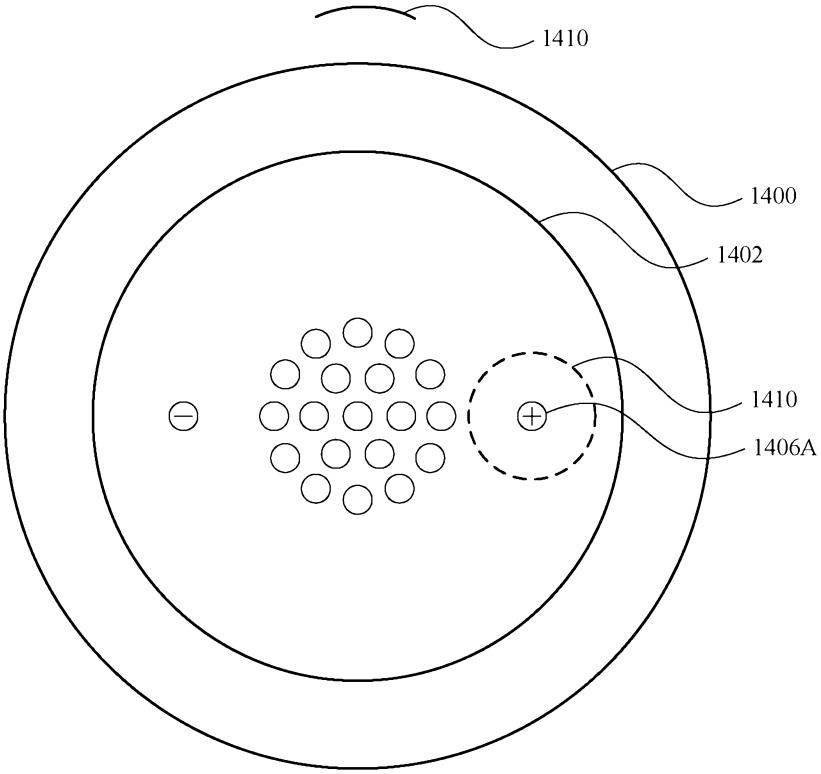


FIG. 14E

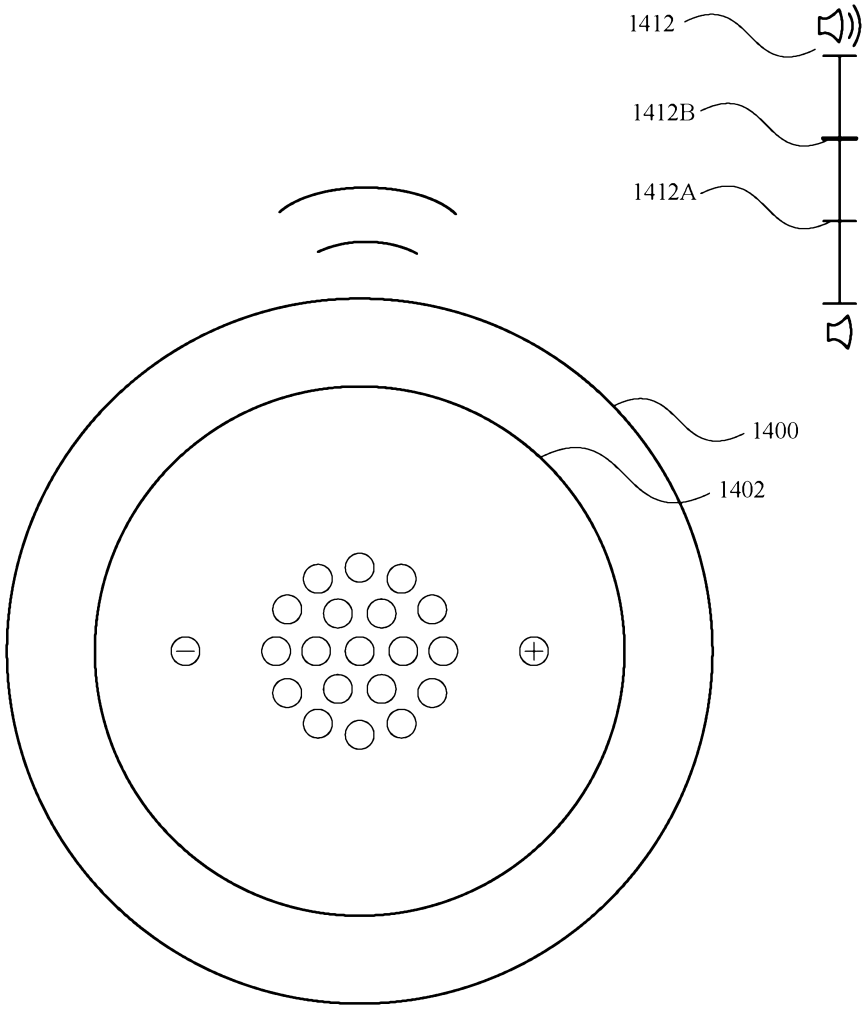


FIG. 14F

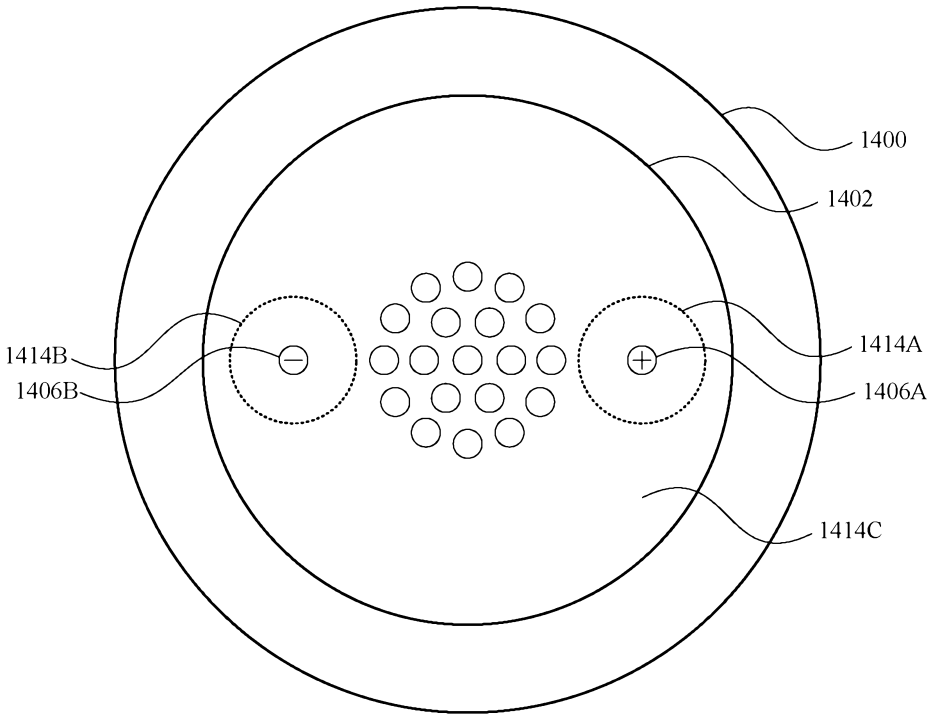


FIG. 14G

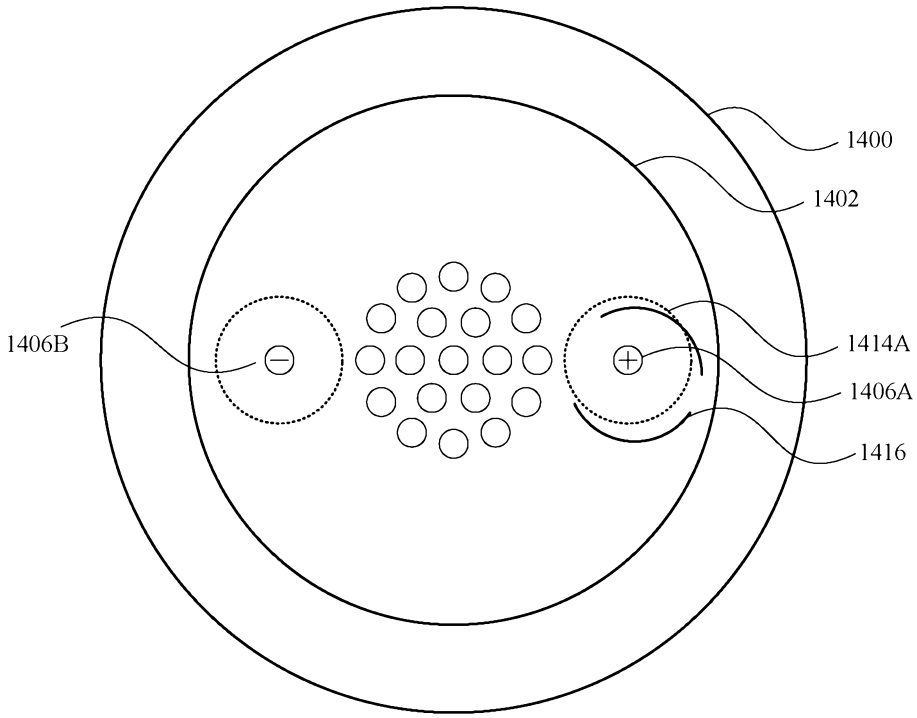


FIG. 14H

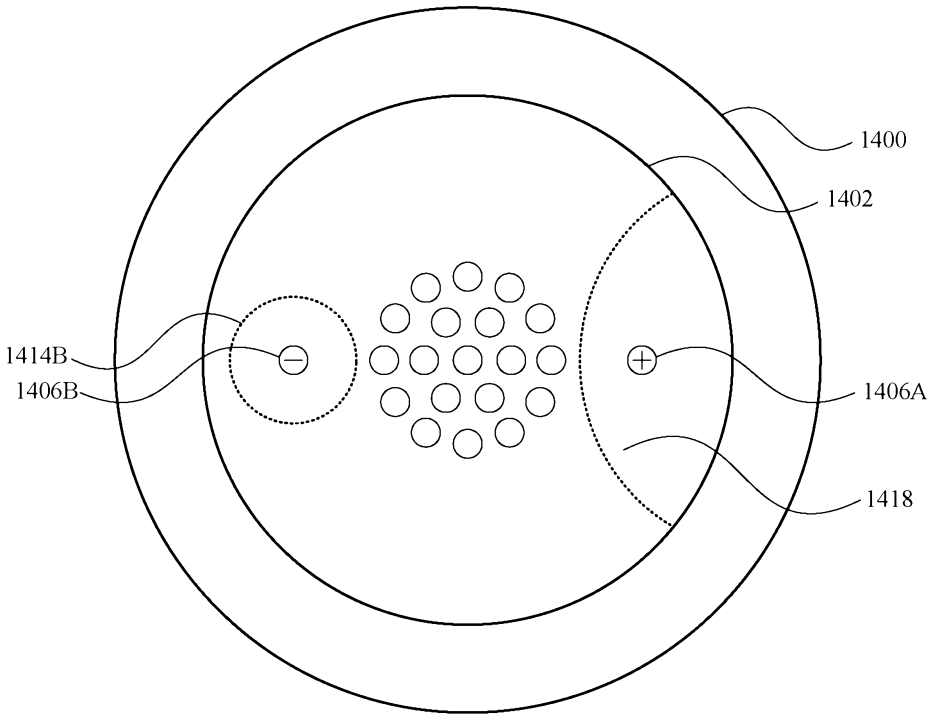


FIG. 14I

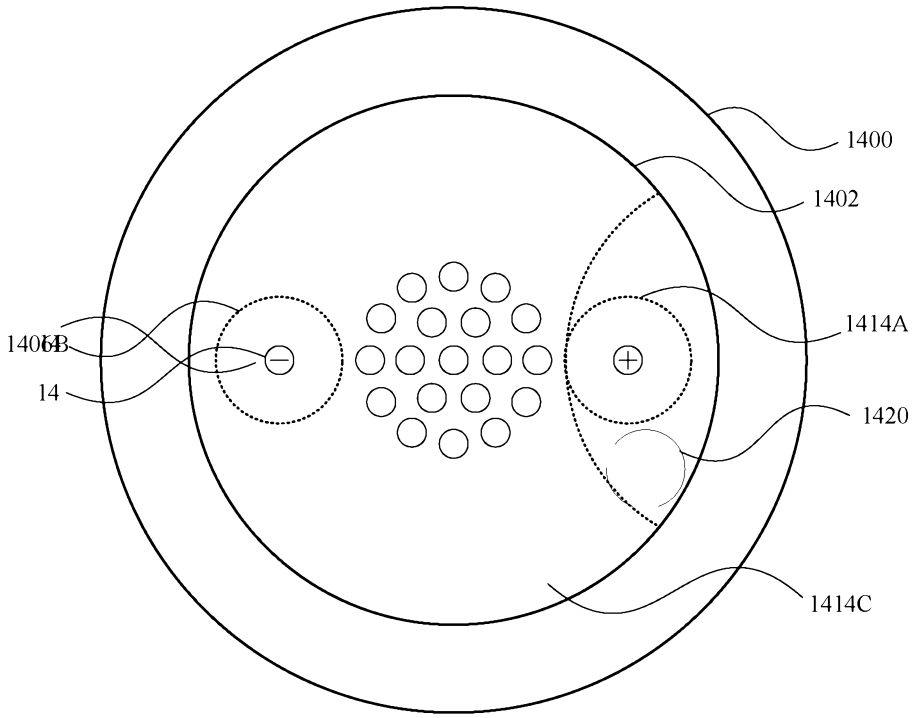


FIG. 14J

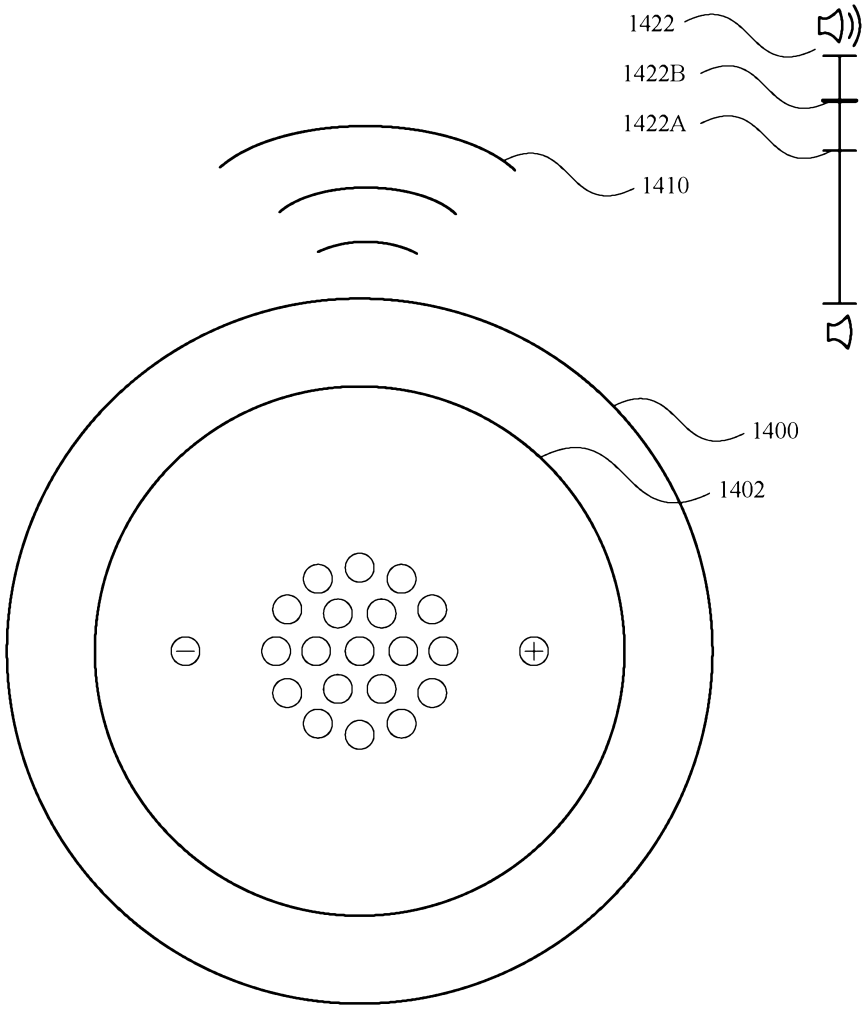


FIG. 14K

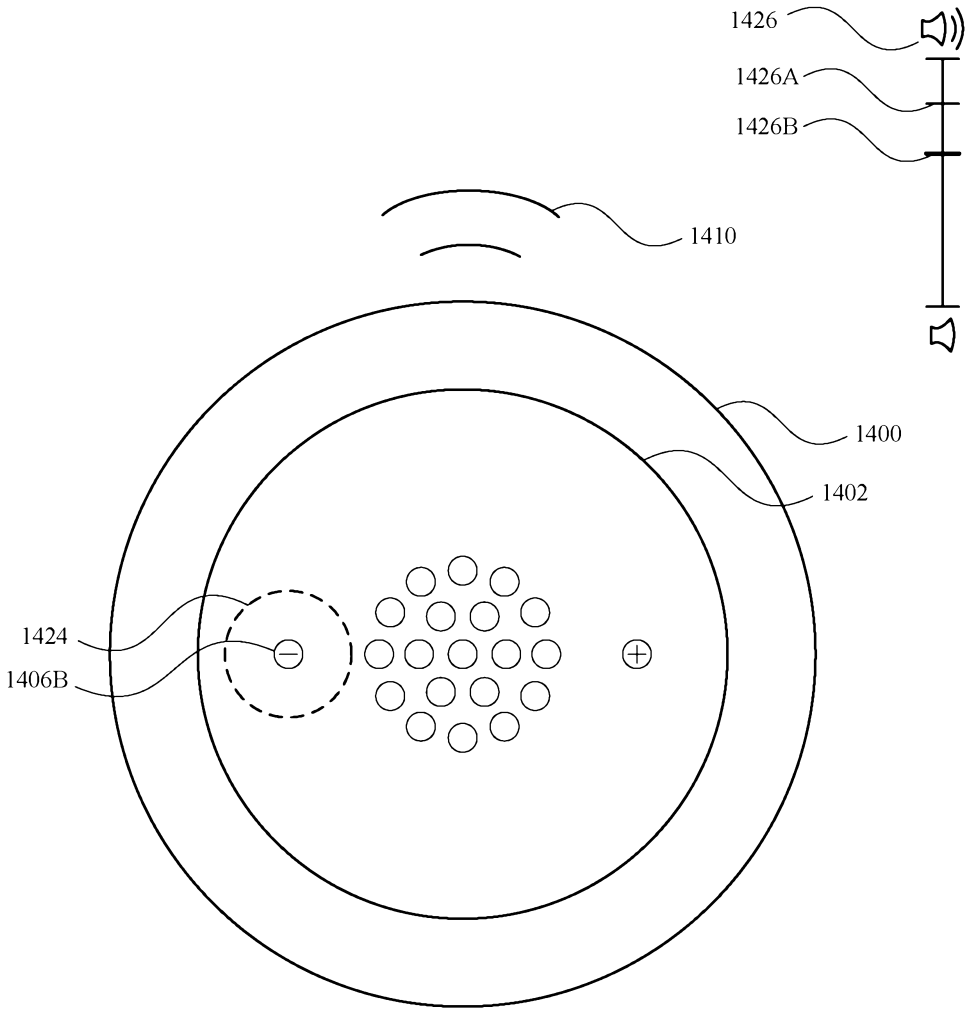


FIG. 14L

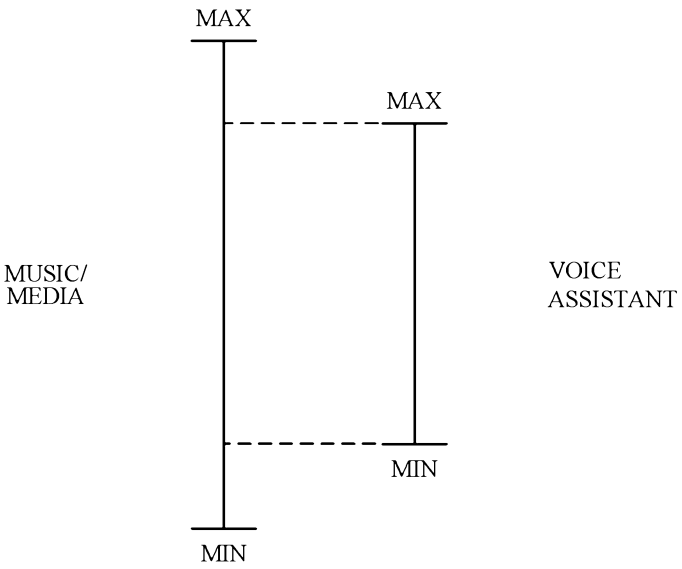


FIG. 14M

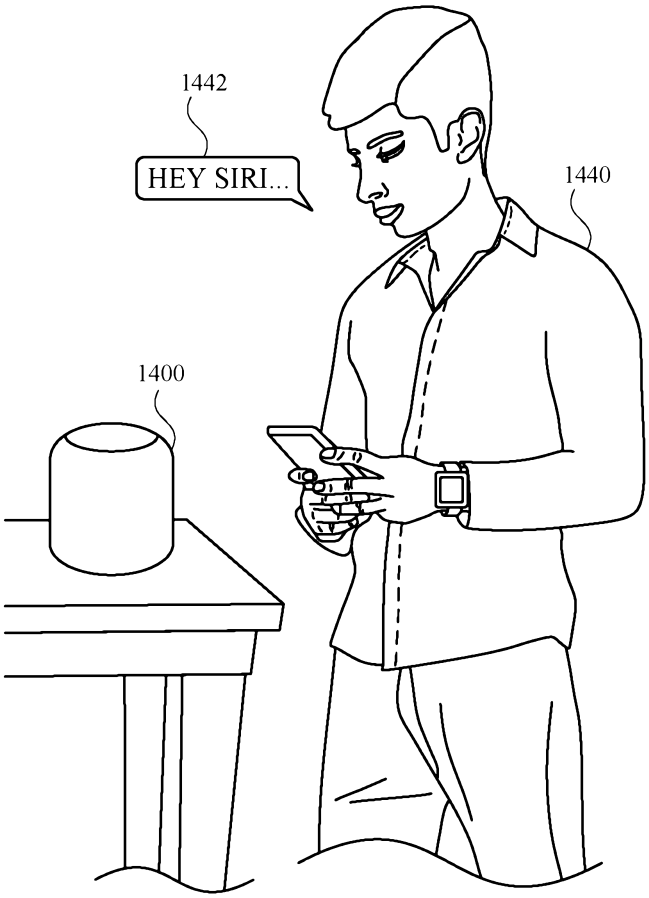


FIG. 14N

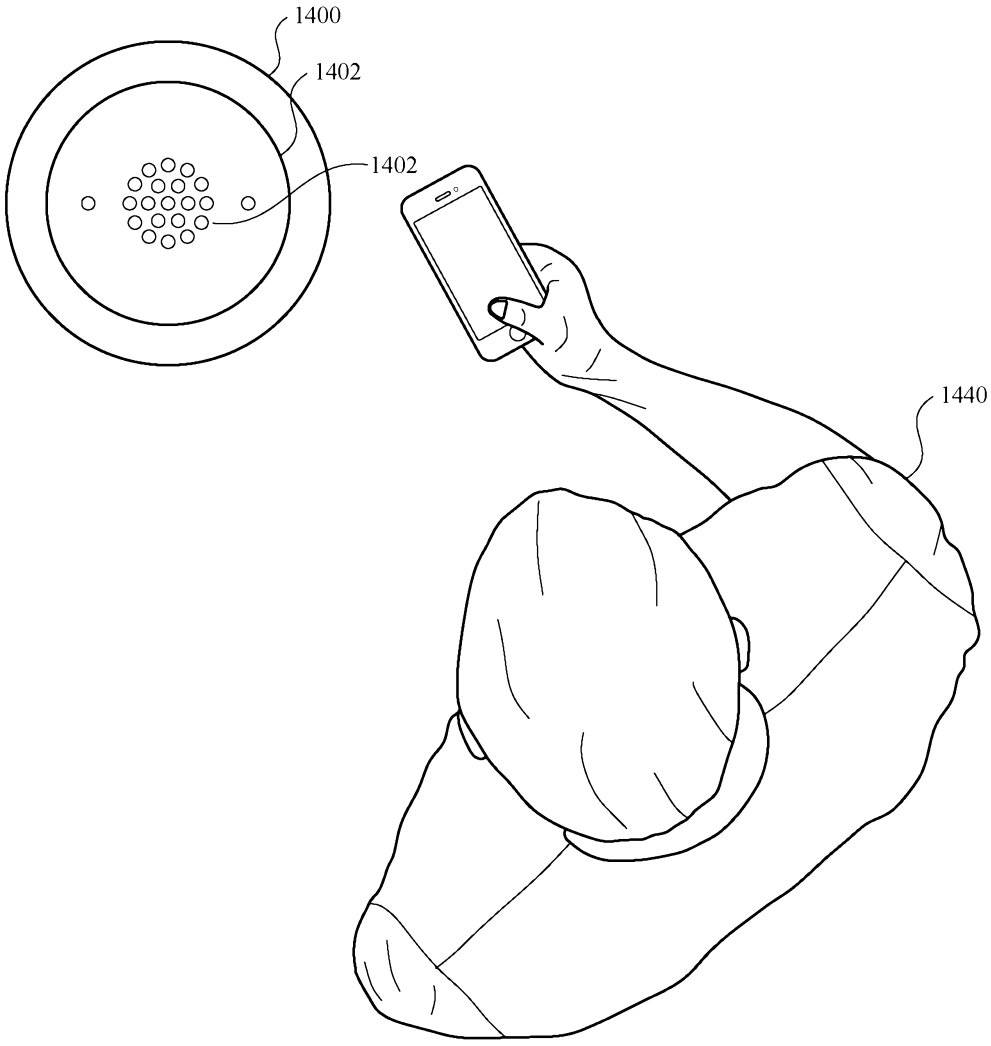


FIG. 14O

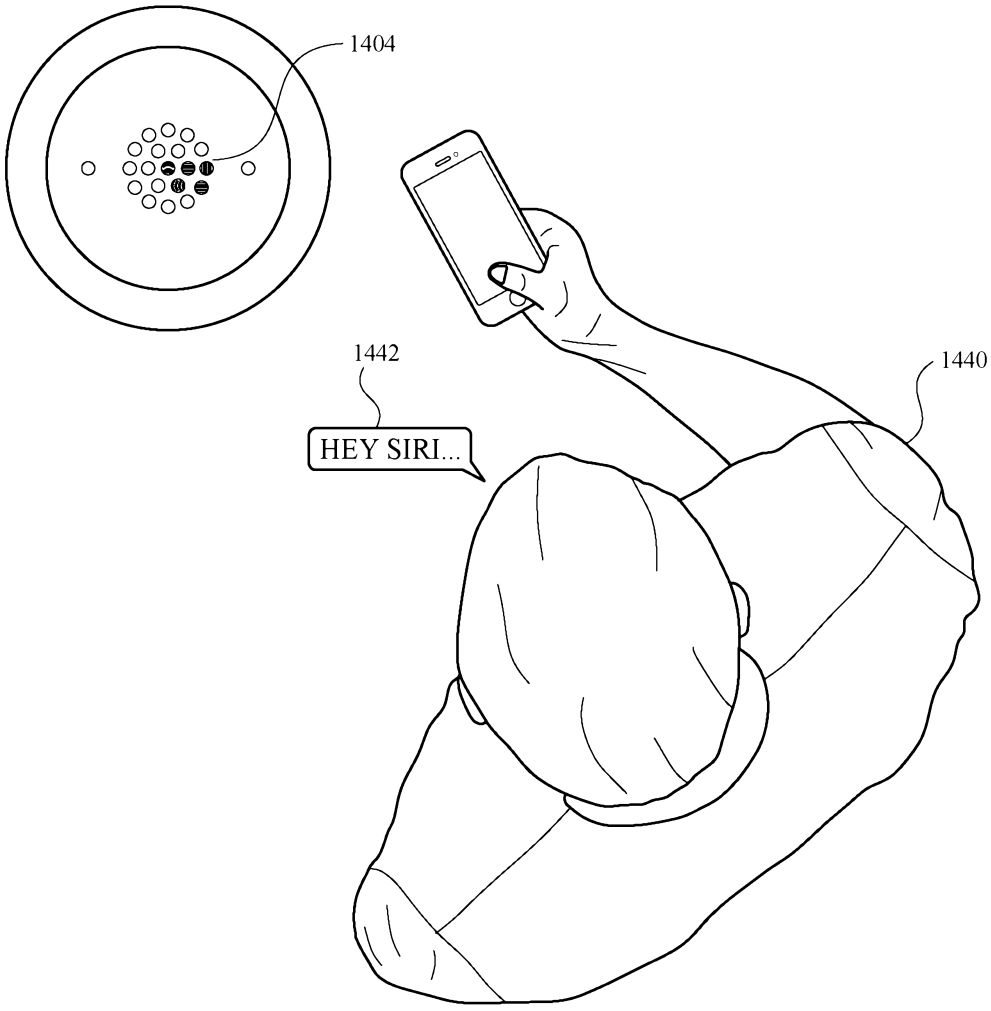


FIG. 14P

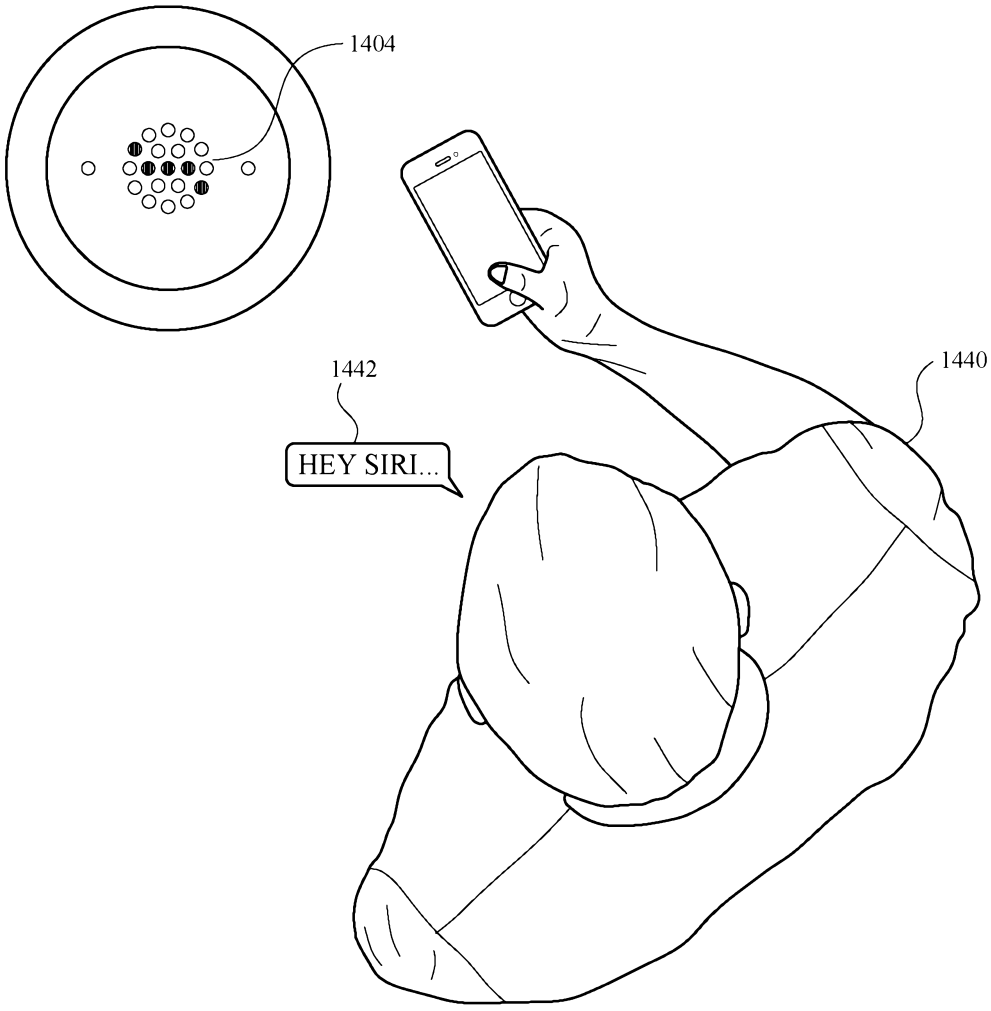


FIG. 14Q

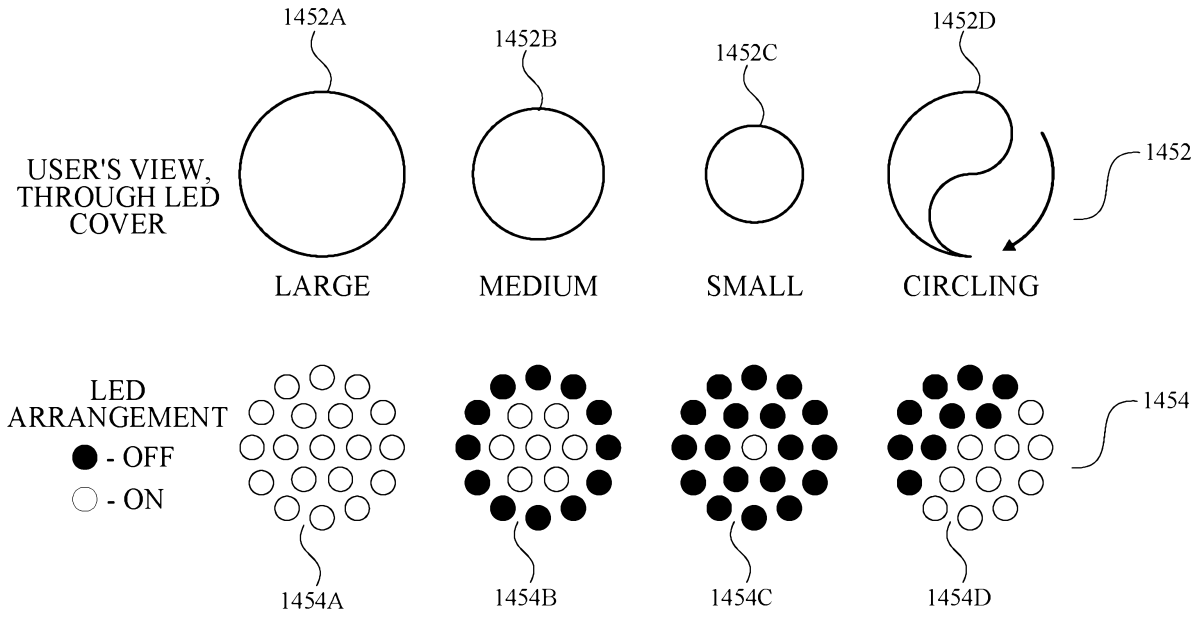

1452A
1452B
1452C
1452D
1452
USER'S VIEW, THROUGH LED COVER
LARGE
MEDIUM
SMALL
CIRCLING
LED ARRANGEMENT
- OFF
- ON
1454
1454A
1454B
1454C
1454D

FIG. 14R

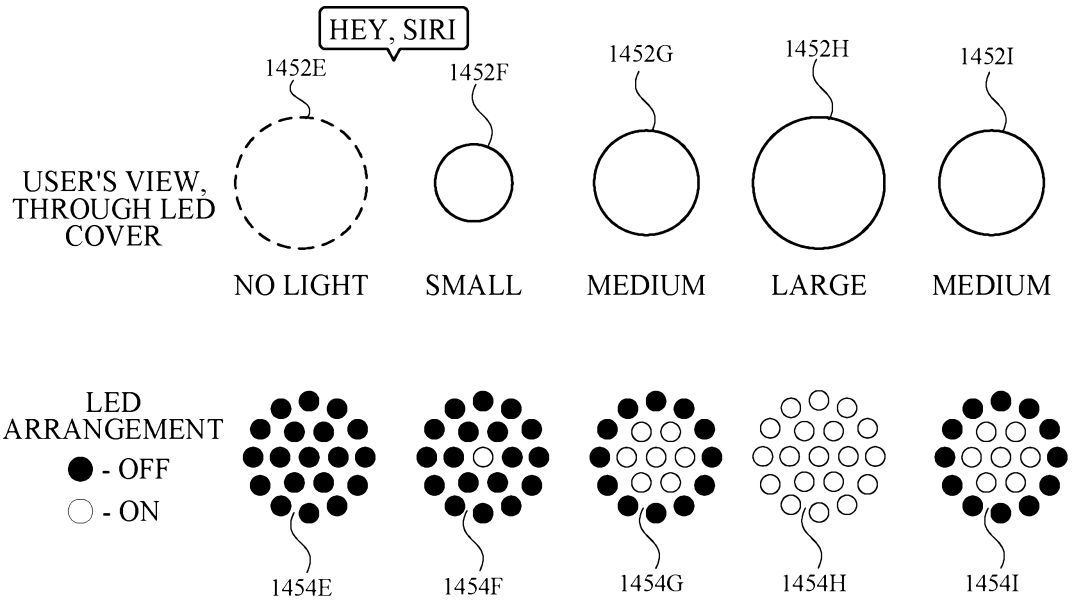

HEY, SIRI
1452E
1452F
1452G
1452H
1452I
USER'S VIEW, THROUGH LED COVER
NO LIGHT
SMALL
MEDIUM
LARGE
MEDIUM
LED ARRANGEMENT
- OFF
- ON
1454E
1454F
1454G
1454H
1454I

FIG. 14S

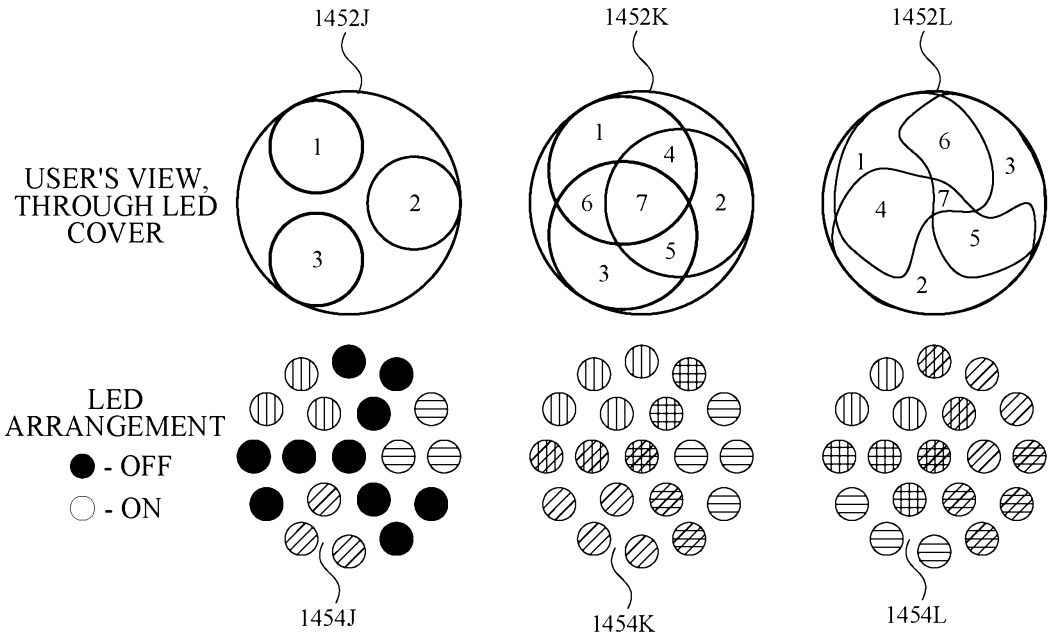

1452J
1452K
1452L
USER'S VIEW, THROUGH LED COVER
1
2
3
1
4
6
7
2
5
3
1
6
3
4
7
5
2
LED ARRANGEMENT
- OFF
- ON
1454J
1454K
1454L

FIG. 14T

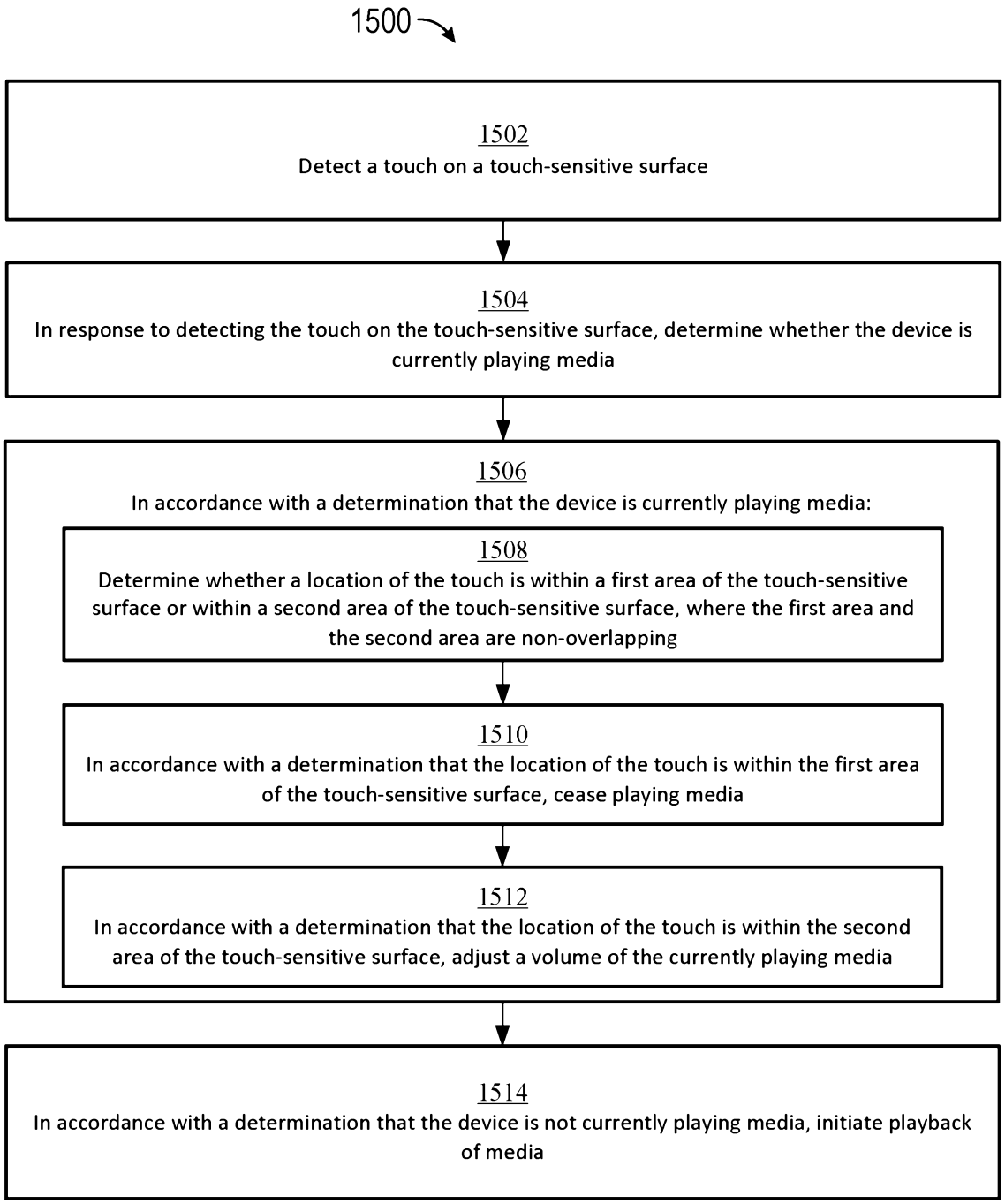

1500
1502
Detect a touch on a touch-sensitive surface
1504
In response to detecting the touch on the touch-sensitive surface, determine whether the device is currently playing media
1506
In accordance with a determination that the device is currently playing media:
1508
Determine whether a location of the touch is within a first area of the touch-sensitive surface or within a second area of the touch-sensitive surface, where the first area and the second area are non-overlapping
1510
In accordance with a determination that the location of the touch is within the first area of the touch-sensitive surface, cease playing media
1512
In accordance with a determination that the location of the touch is within the second area of the touch-sensitive surface, adjust a volume of the currently playing media
1514
In accordance with a determination that the device is not currently playing media, initiate playback of media

FIG. 15

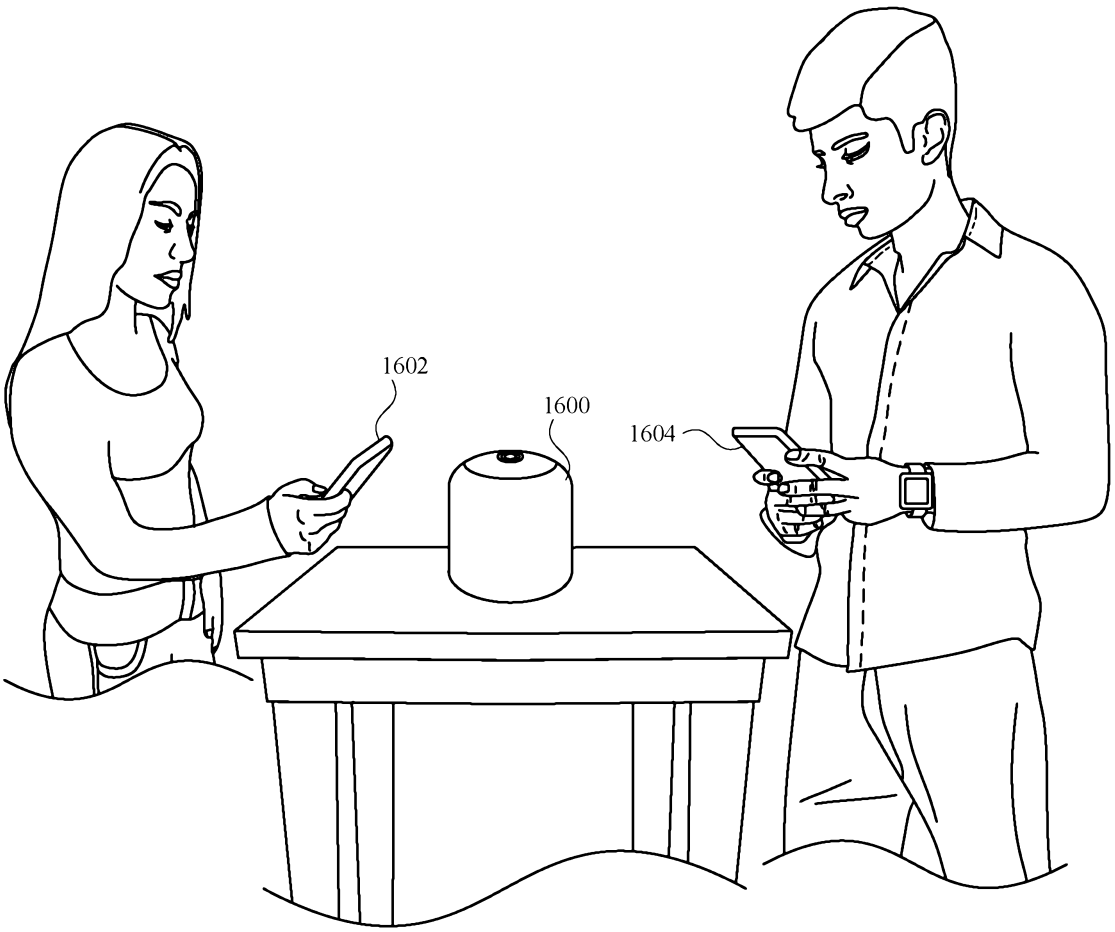


FIG. 16A

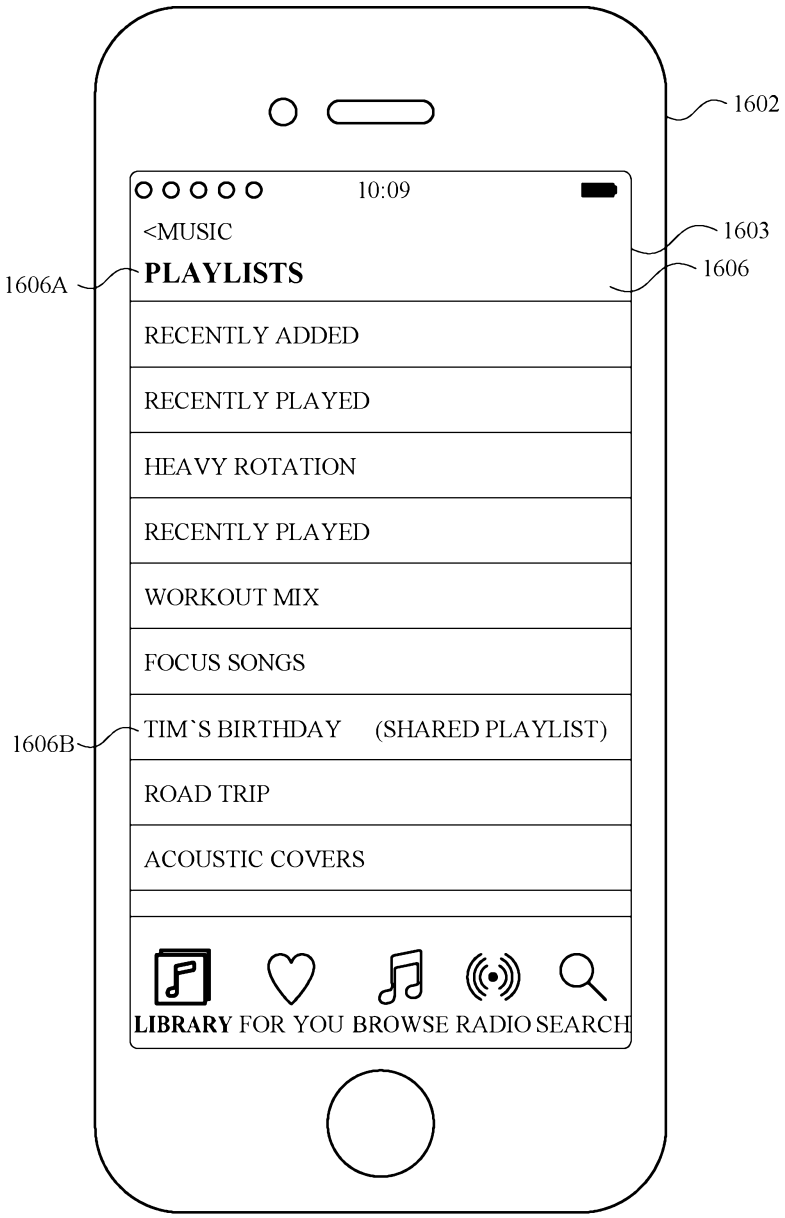

1602
10:09
<MUSIC
1603
1606A
PLAYLISTS
1606
RECENTLY ADDED
RECENTLY PLAYED
HEAVY ROTATION
RECENTLY PLAYED
WORKOUT MIX
FOCUS SONGS
1606B
TIM`S BIRTHDAY (SHARED PLAYLIST)
ROAD TRIP
ACOUSTIC COVERS
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16B

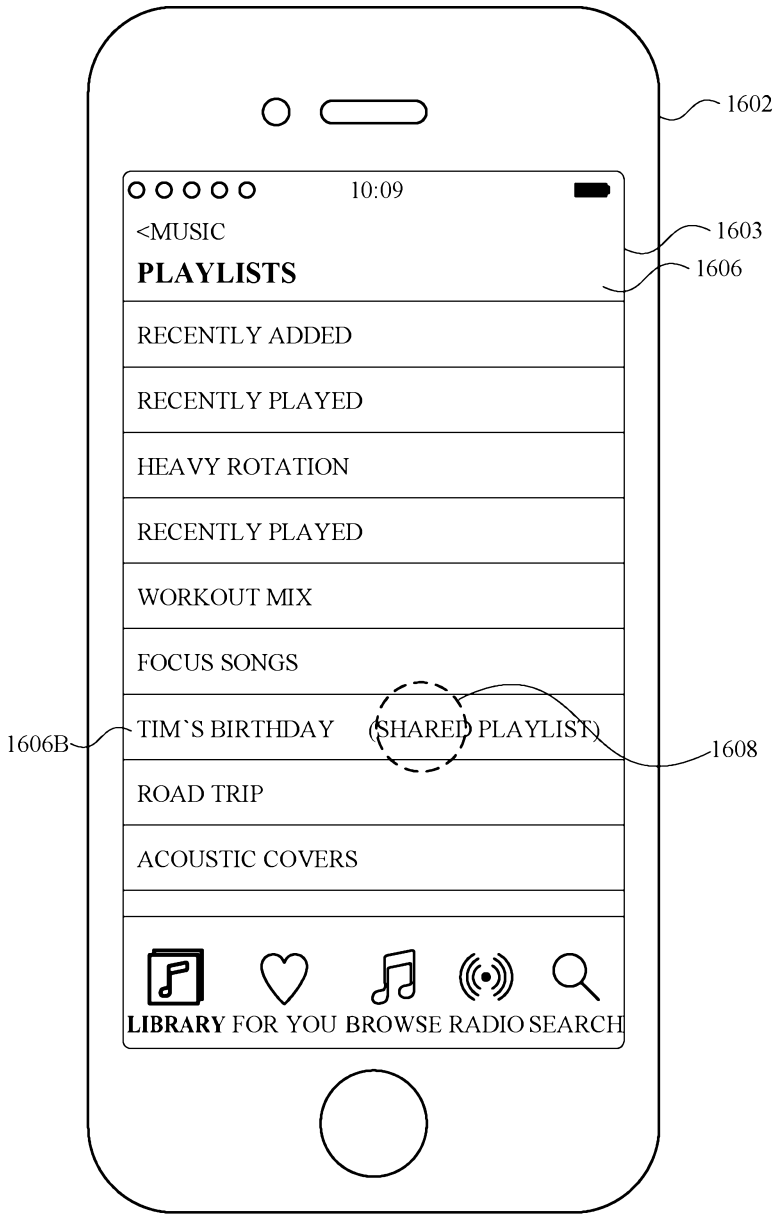

1602
10:09
<MUSIC
PLAYLISTS
1603
1606
RECENTLY ADDED
RECENTLY PLAYED
HEAVY ROTATION
RECENTLY PLAYED
WORKOUT MIX
FOCUS SONGS
TIM`S BIRTHDAY (SHARED PLAYLIST)
1606B
1608
ROAD TRIP
ACOUSTIC COVERS
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16C

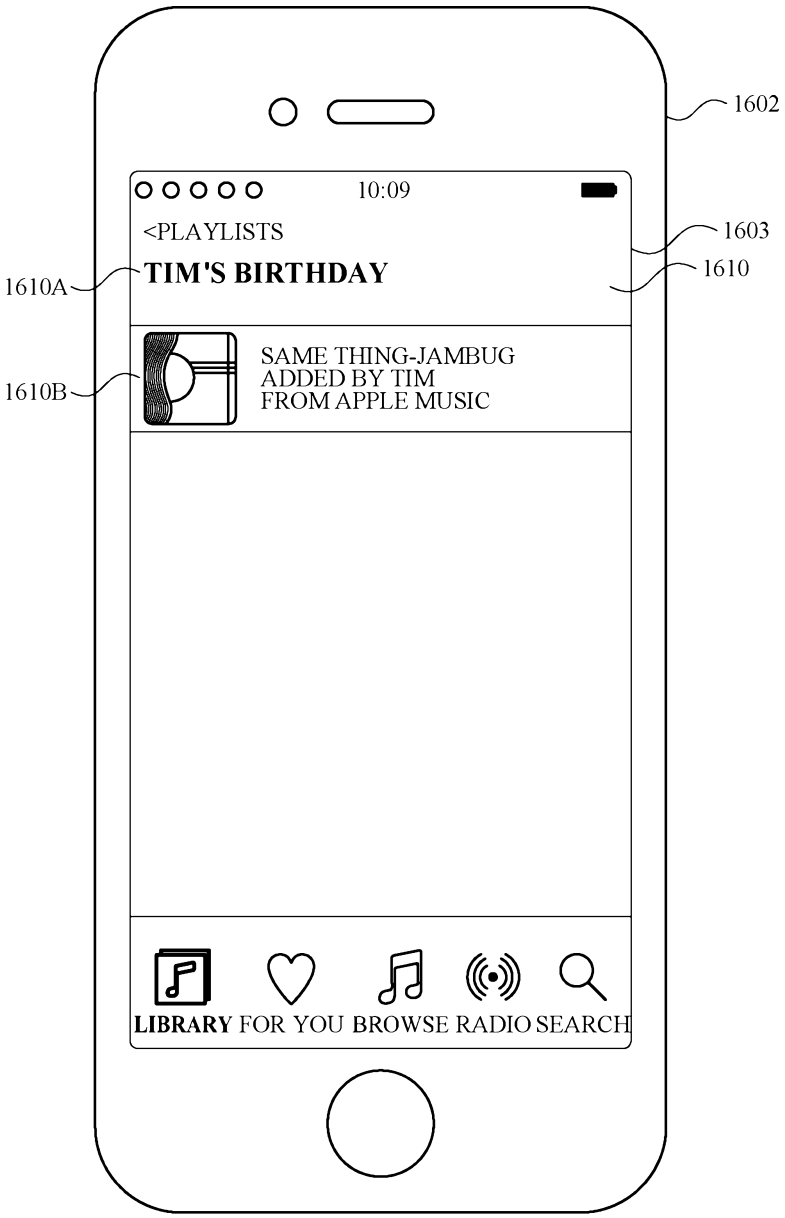

1602
10:09
<PLAYLISTS
1603
1610A
TIM'S BIRTHDAY
1610
1610B
SAME THING-JAMBUG
ADDED BY TIM
FROM APPLE MUSIC
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16D

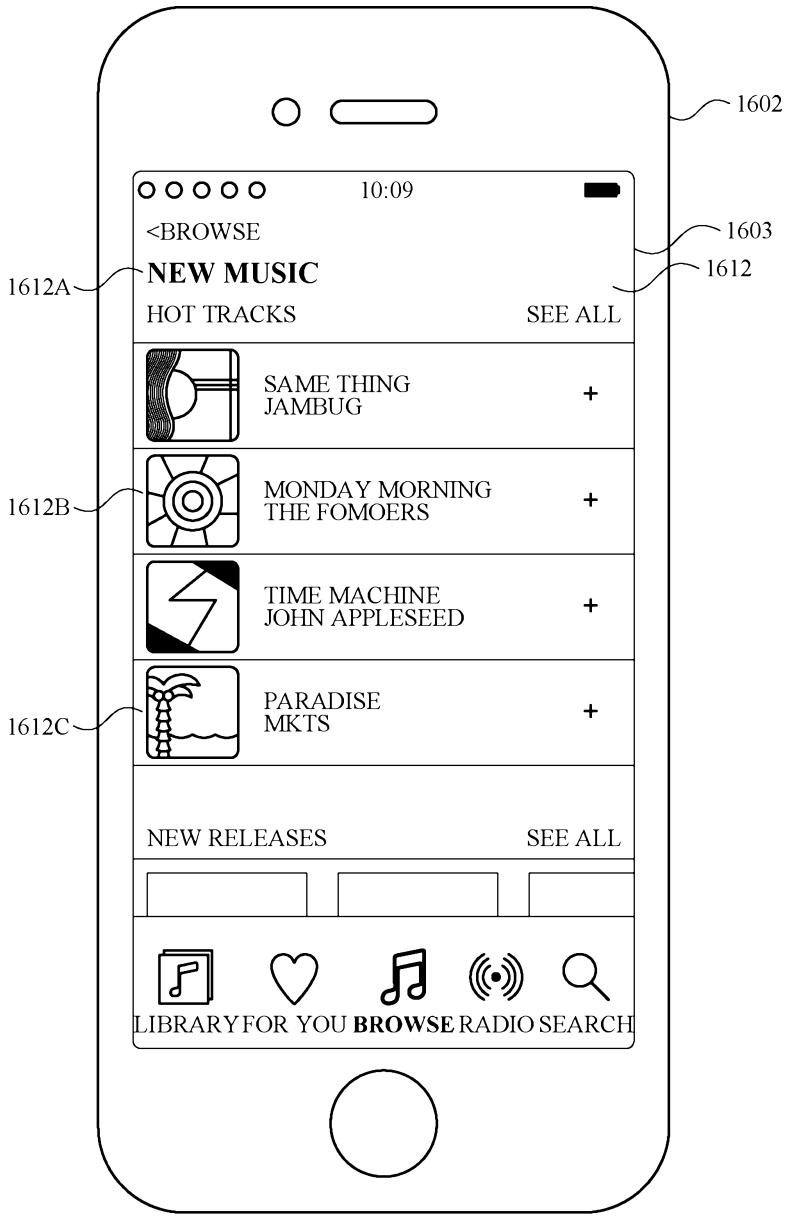

1602
10:09
<BROWSE
1603
1612A
NEW MUSIC
1612
HOT TRACKS
SEE ALL
SAME THING
JAMBUG
+
1612B
MONDAY MORNING
THE FOMOERS
+
TIME MACHINE
JOHN APPLESEED
+
1612C
PARADISE
MKTS
+
NEW RELEASES
SEE ALL
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16E

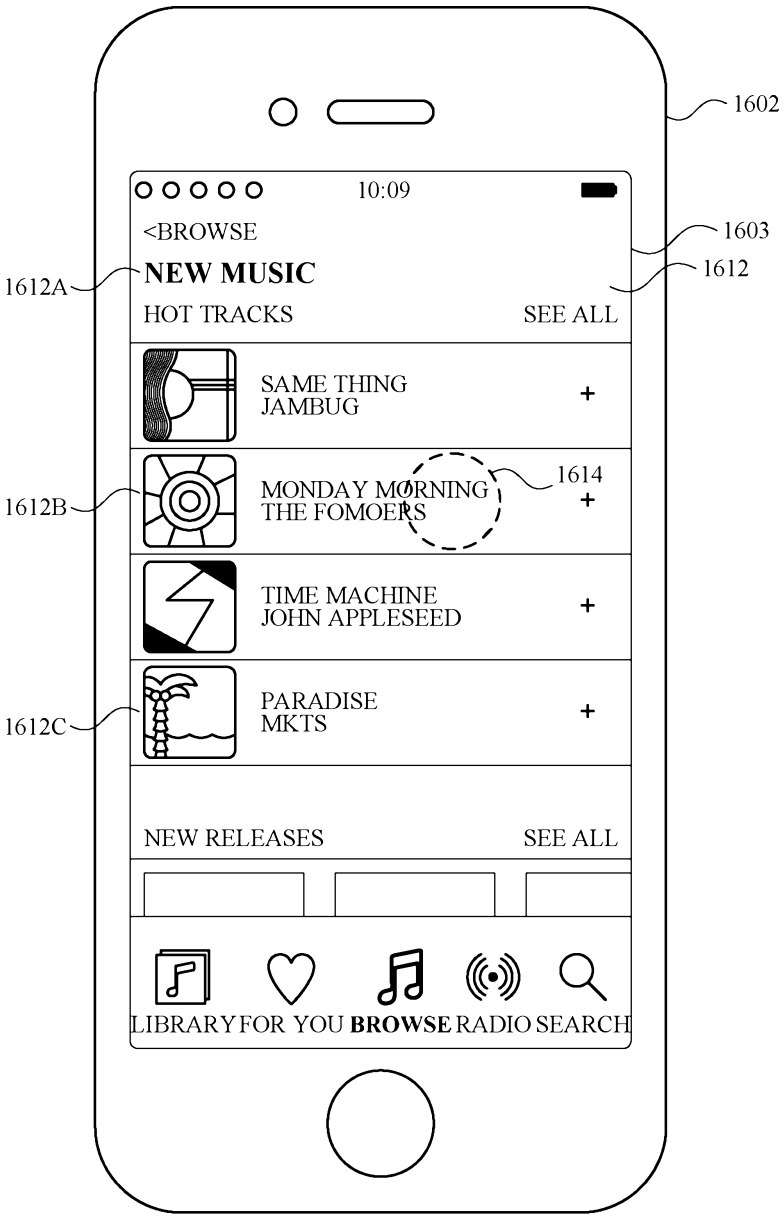

1602
10:09
<BROWSE
1603
NEW MUSIC
1612
1612A
HOT TRACKS
SEE ALL
SAME THING
JAMBUG
+
MONDAY MORNING
THE FOMOERS
1614
1612B
+
TIME MACHINE
JOHN APPLESEED
+
PARADISE
MKTS
1612C
+
NEW RELEASES
SEE ALL
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16F

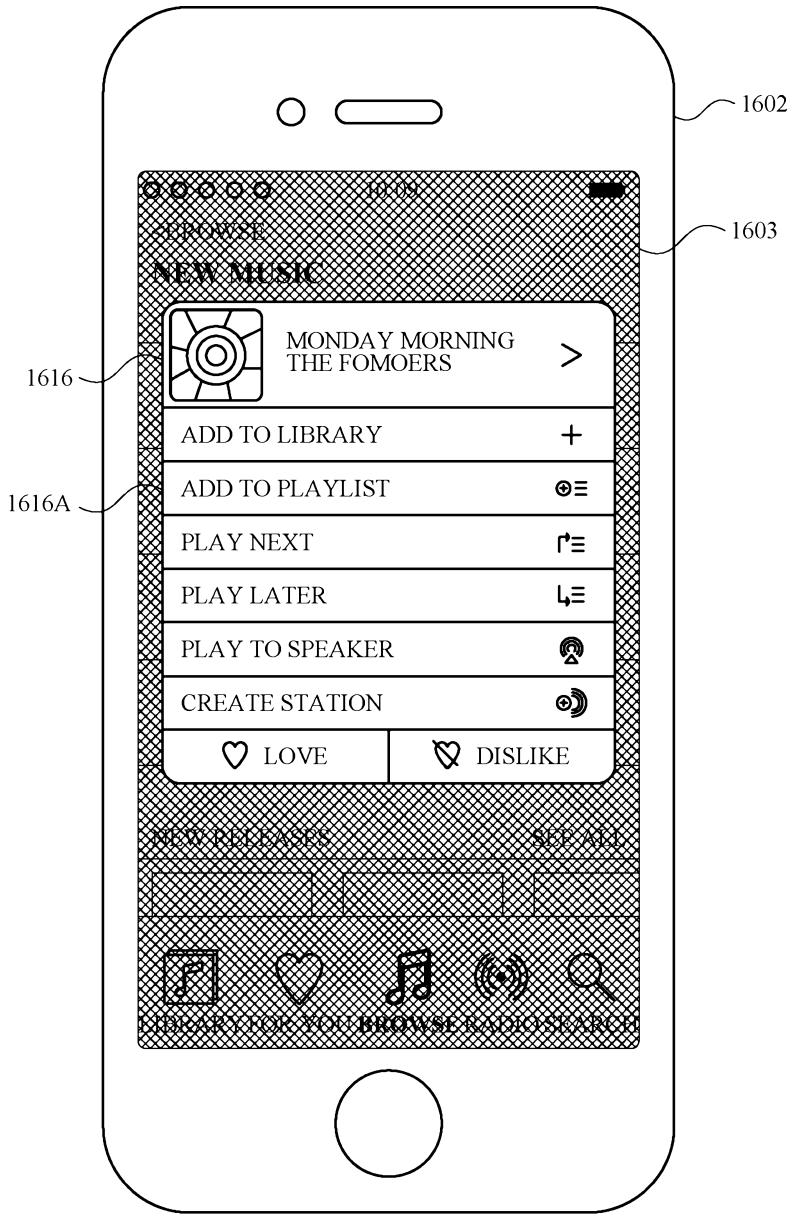

1602
1603
1616
1616A
10:09
BROWSE
NEW MUSIC
MONDAY MORNING
THE FOMOERS
ADD TO LIBRARY
ADD TO PLAYLIST
PLAY NEXT
PLAY LATER
PLAY TO SPEAKER
CREATE STATION
LOVE
DISLIKE
NEW RELEASES
SEE ALL
LIBRARY
FOR YOU
BROWSE
RADIO
SEARCH

FIG. 16G

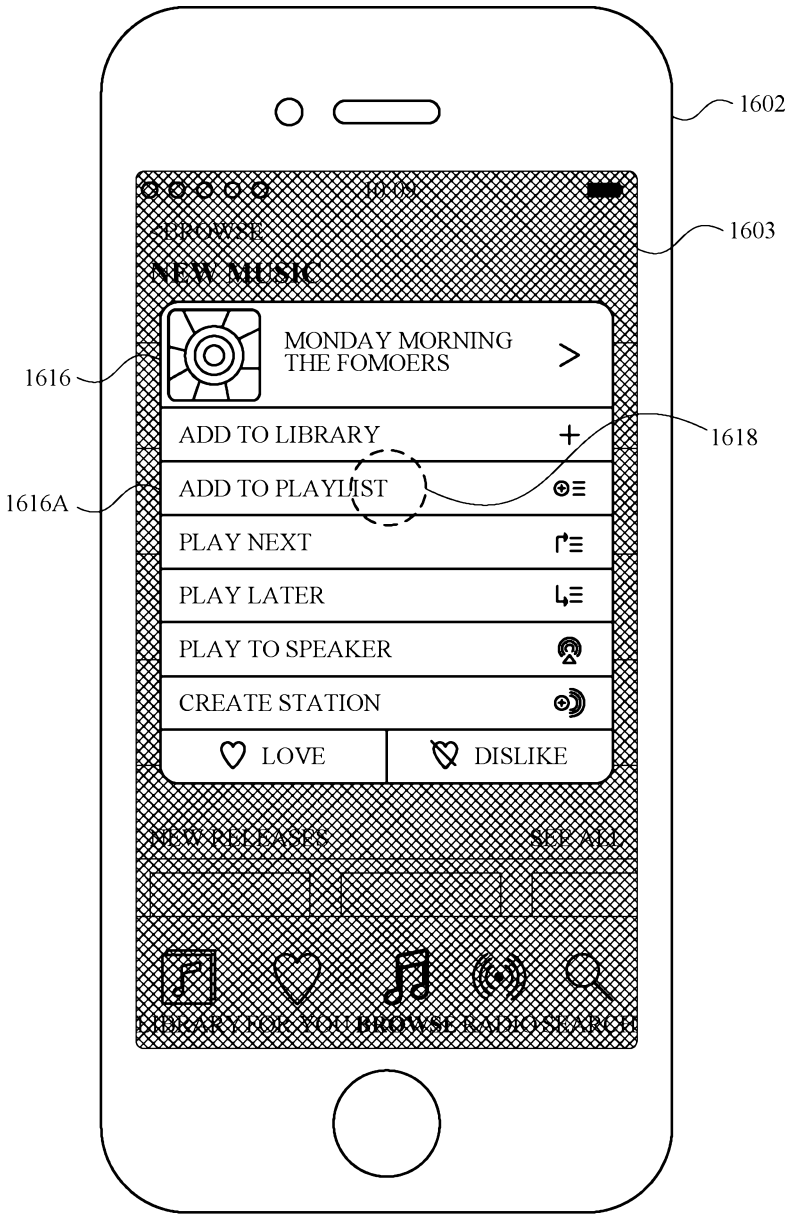

1602
1603
1616
1616A
1618
10:09
BROWSE
NEW MUSIC
MONDAY MORNING
THE FOMOERS
ADD TO LIBRARY
ADD TO PLAYLIST
PLAY NEXT
PLAY LATER
PLAY TO SPEAKER
CREATE STATION
LOVE
DISLIKE
NEW RELEASES
SEE ALL
LIBRARY
FOR YOU
BROWSE
RADIO
SEARCH

FIG. 16H

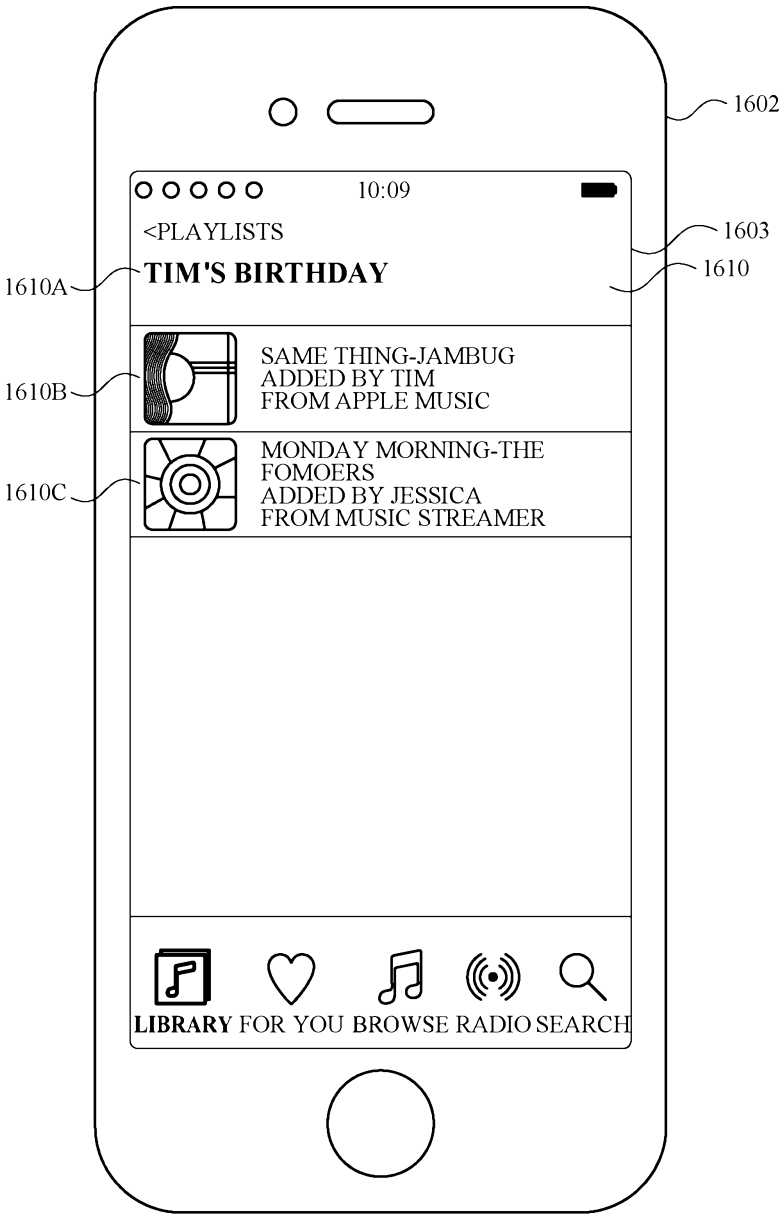

1602
10:09
<PLAYLISTS
1603
TIM'S BIRTHDAY
1610
1610A
SAME THING-JAMBUG
ADDED BY TIM
FROM APPLE MUSIC
1610B
MONDAY MORNING-THE
FOMOERS
ADDED BY JESSICA
FROM MUSIC STREAMER
1610C
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16I

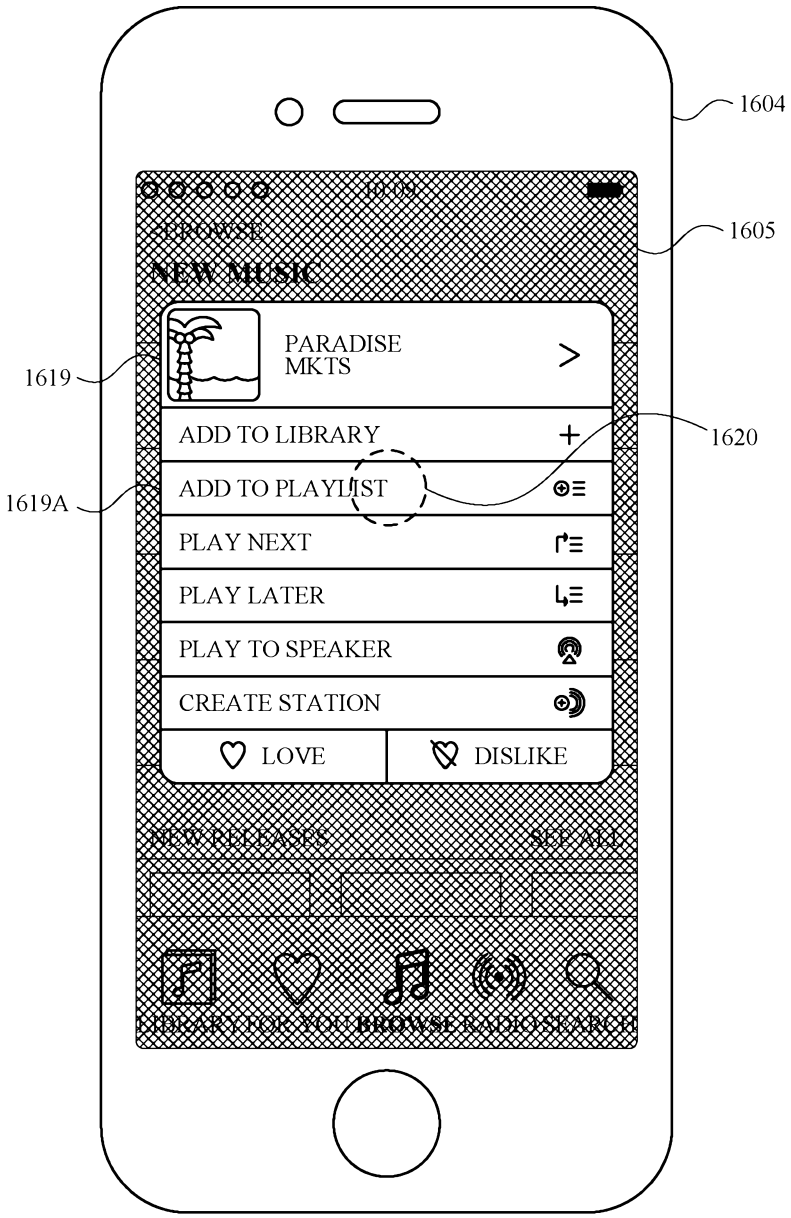

1604
1605
1619
1619A
1620
10:09
BROWSE
NEW MUSIC
PARADISE
MKTS
ADD TO LIBRARY
ADD TO PLAYLIST
PLAY NEXT
PLAY LATER
PLAY TO SPEAKER
CREATE STATION
LOVE
DISLIKE
NEW RELEASES
SEE ALL
FOR YOU
BROWSE
RADIO
SEARCH

FIG. 16J

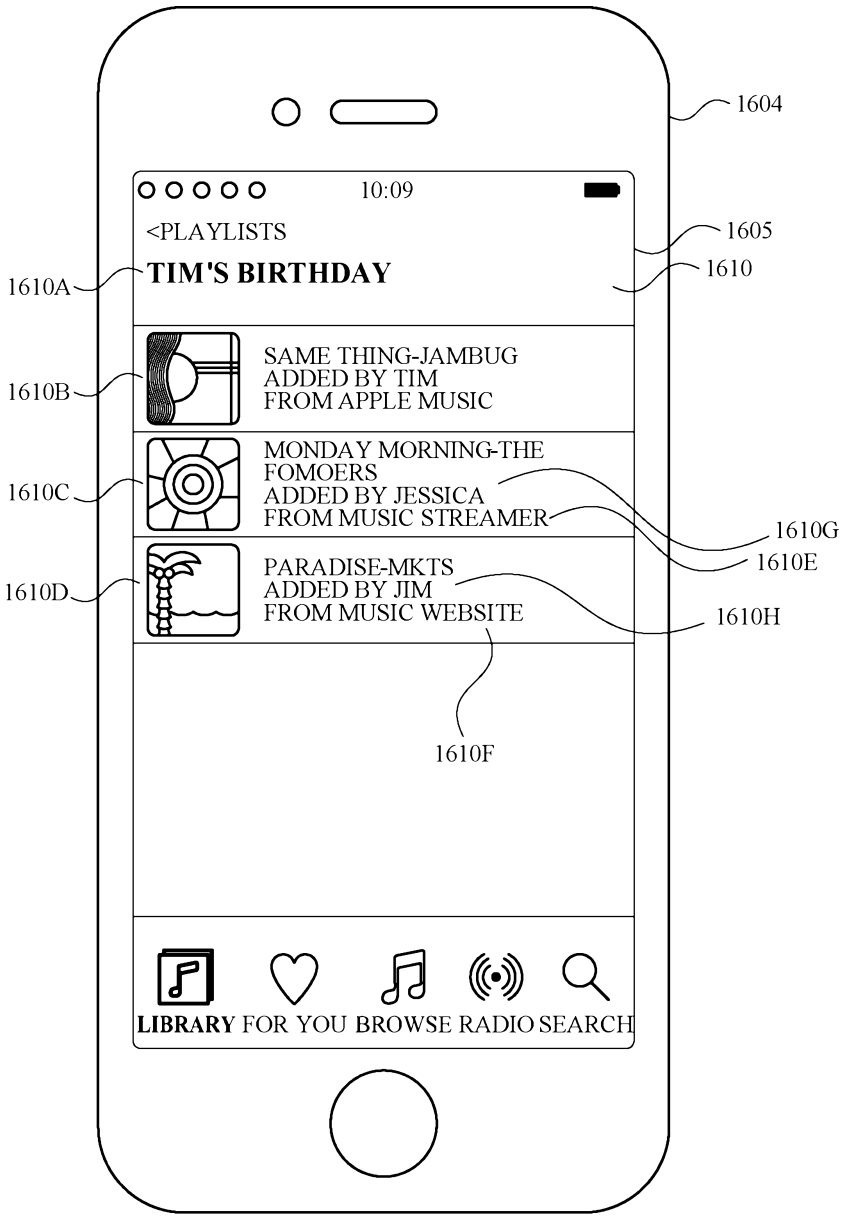

1604
10:09
<PLAYLISTS
1605
TIM'S BIRTHDAY
1610
1610A
SAME THING-JAMBUG
ADDED BY TIM
FROM APPLE MUSIC
1610B
MONDAY MORNING-THE
FOMOERS
ADDED BY JESSICA
FROM MUSIC STREAMER
1610C
1610G
1610E
PARADISE-MKTS
ADDED BY JIM
FROM MUSIC WEBSITE
1610D
1610H
1610F
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16K

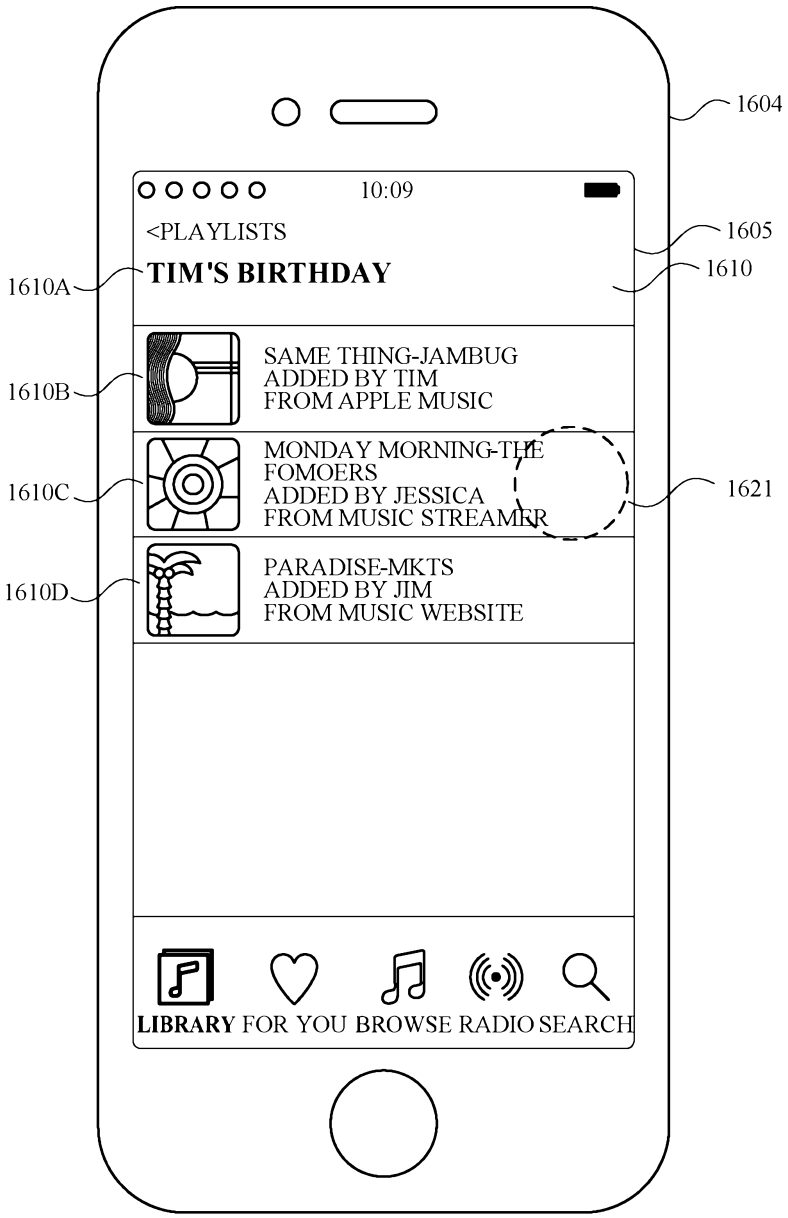

1604
10:09
<PLAYLISTS
1605
TIM'S BIRTHDAY
1610
1610A
SAME THING-JAMBUG
ADDED BY TIM
FROM APPLE MUSIC
1610B
MONDAY MORNING-THE
FOMOERS
ADDED BY JESSICA
FROM MUSIC STREAMER
1610C
1621
PARADISE-MKTS
ADDED BY JIM
FROM MUSIC WEBSITE
1610D
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16L

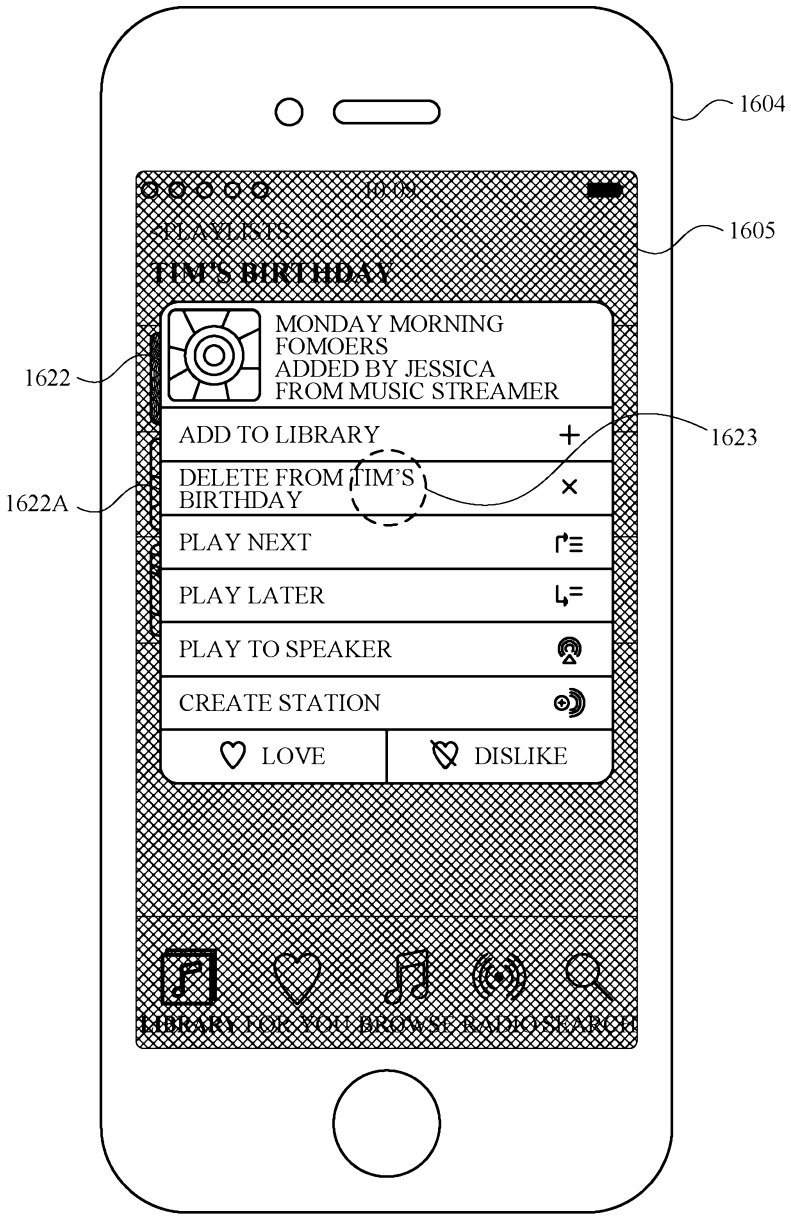

1604
10:09
PLAYLISTS
TIM'S BIRTHDAY
1605
MONDAY MORNING
FOMOERS
ADDED BY JESSICA
FROM MUSIC STREAMER
1622
ADD TO LIBRARY
1623
DELETE FROM TIM'S BIRTHDAY
1622A
PLAY NEXT
PLAY LATER
PLAY TO SPEAKER
CREATE STATION
LOVE
DISLIKE
LIBRARY
FOR YOU
BROWSE
RADIO
SEARCH

FIG. 16M

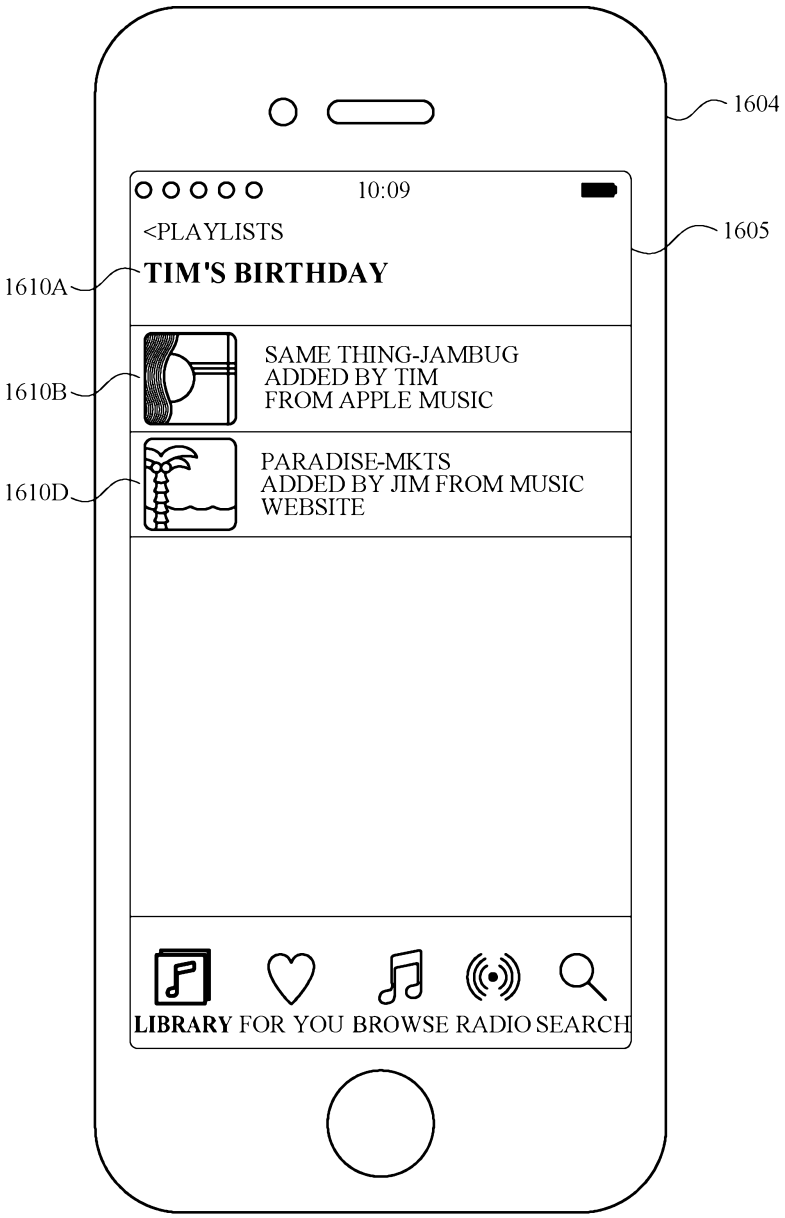

1604
10:09
<PLAYLISTS
1605
1610A
TIM'S BIRTHDAY
1610B
SAME THING-JAMBUG
ADDED BY TIM
FROM APPLE MUSIC
1610D
PARADISE-MKTS
ADDED BY JIM FROM MUSIC
WEBSITE
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16N

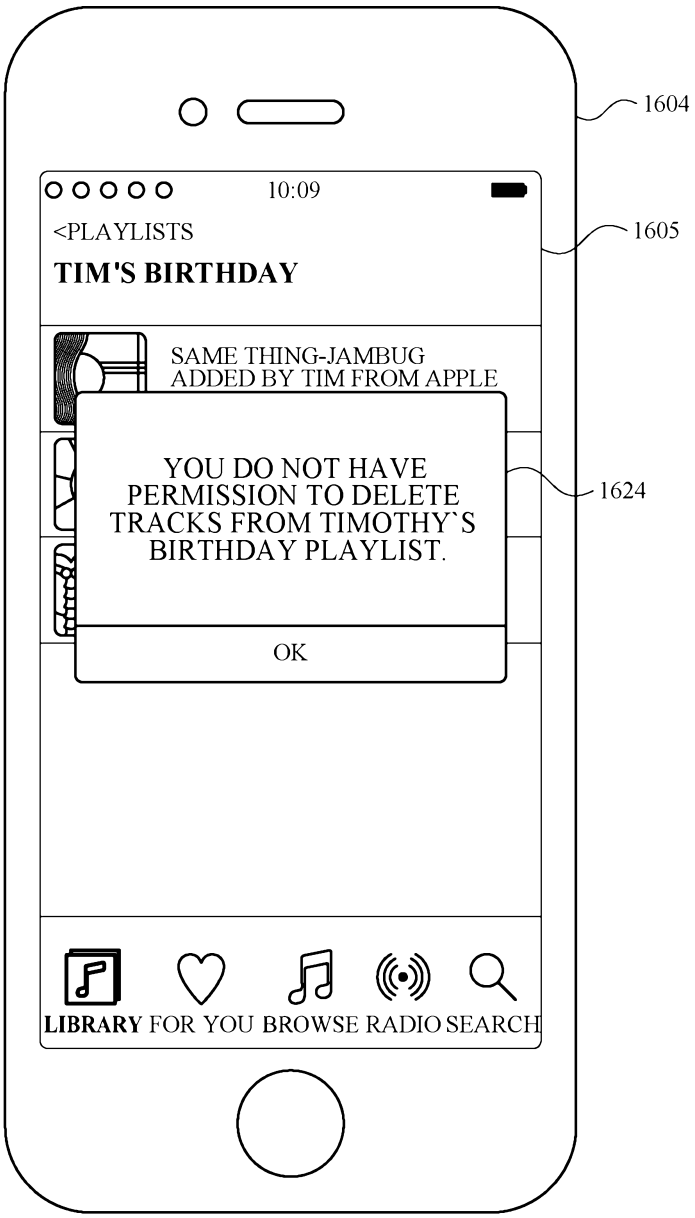

1604
10:09
<PLAYLISTS
1605
TIM'S BIRTHDAY
SAME THING-JAMBUG
ADDED BY TIM FROM APPLE
YOU DO NOT HAVE PERMISSION TO DELETE TRACKS FROM TIMOTHY`S BIRTHDAY PLAYLIST.
1624
OK
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16O

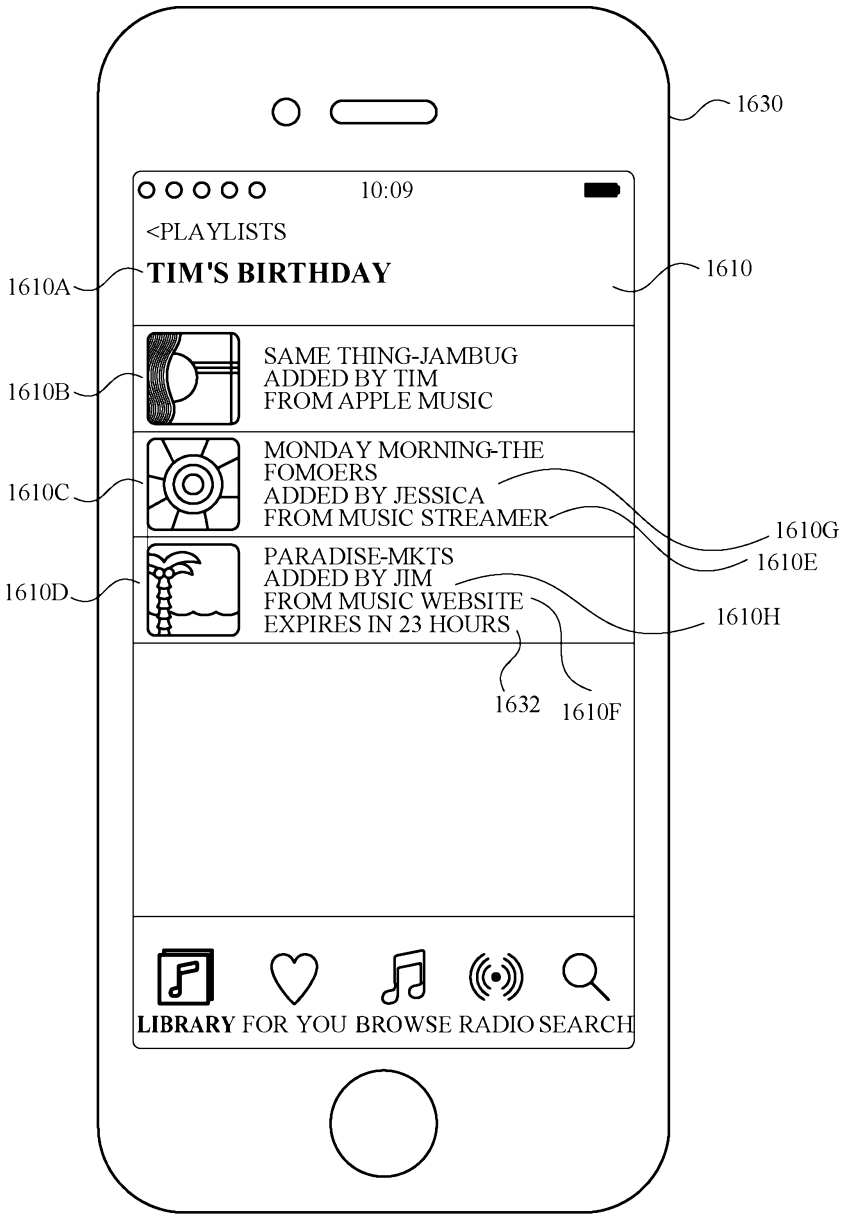

1630
10:09
<PLAYLISTS
TIM'S BIRTHDAY
1610A
1610
1610B
SAME THING-JAMBUG
ADDED BY TIM
FROM APPLE MUSIC
1610C
MONDAY MORNING-THE
FOMOERS
ADDED BY JESSICA
FROM MUSIC STREAMER
1610G
1610E
1610D
PARADISE-MKTS
ADDED BY JIM
FROM MUSIC WEBSITE
EXPIRES IN 23 HOURS
1610H
1632
1610F
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 16P

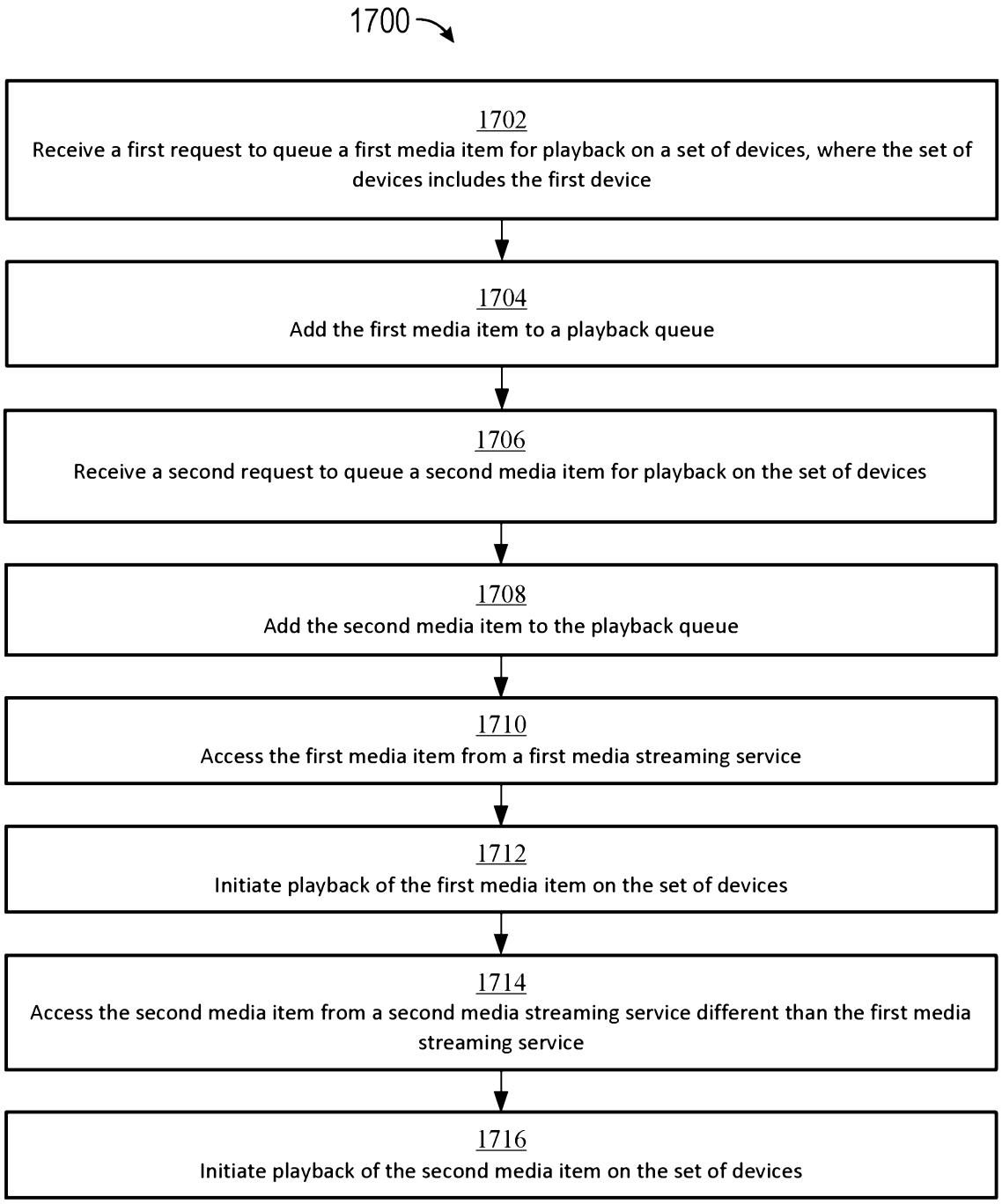

1700
1702
Receive a first request to queue a first media item for playback on a set of devices, where the set of devices includes the first device
1704
Add the first media item to a playback queue
1706
Receive a second request to queue a second media item for playback on the set of devices
1708
Add the second media item to the playback queue
1710
Access the first media item from a first media streaming service
1712
Initiate playback of the first media item on the set of devices
1714
Access the second media item from a second media streaming service different than the first media streaming service
1716
Initiate playback of the second media item on the set of devices

FIG. 17

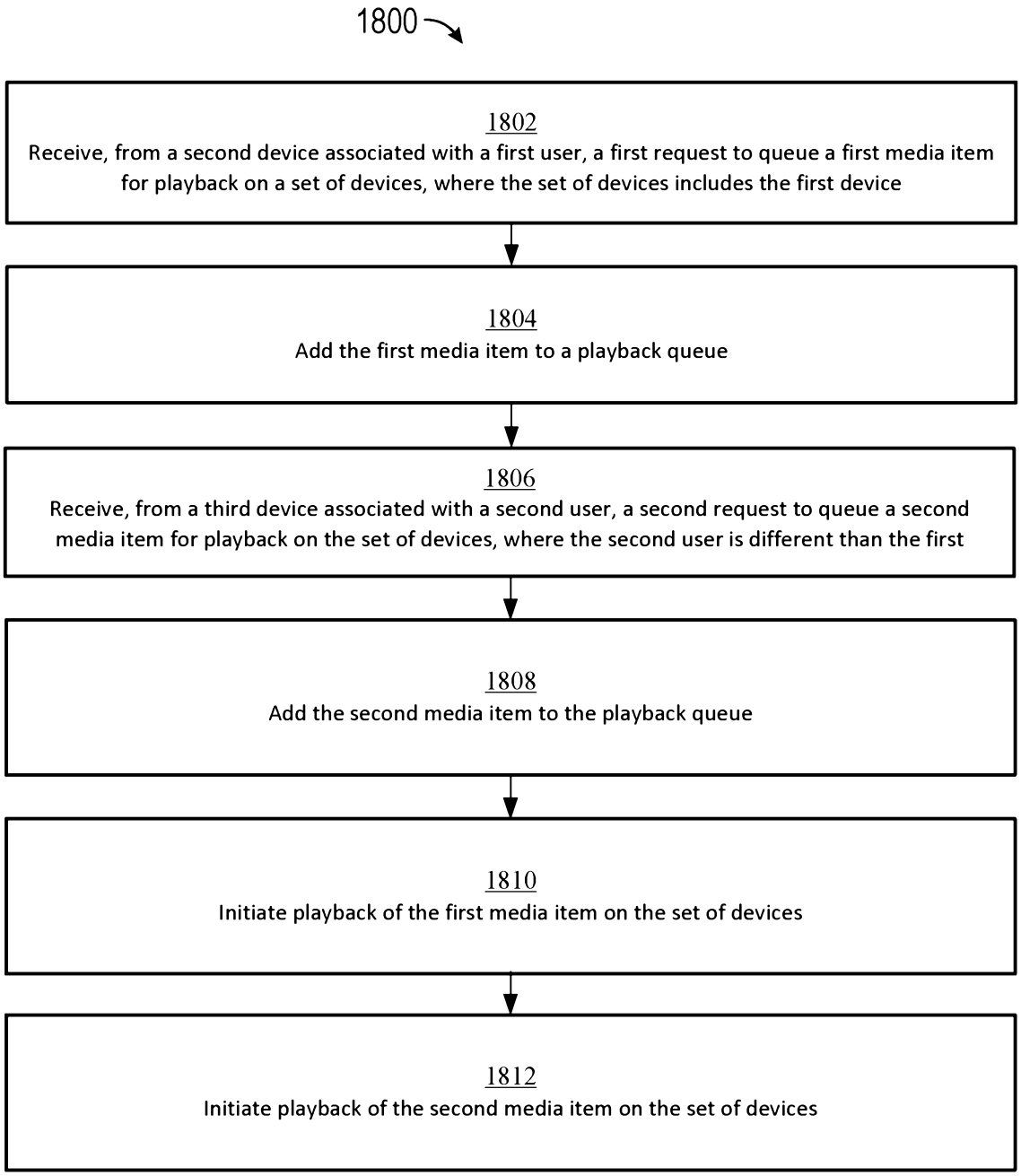

1800
1802
Receive, from a second device associated with a first user, a first request to queue a first media item for playback on a set of devices, where the set of devices includes the first device
1804
Add the first media item to a playback queue
1806
Receive, from a third device associated with a second user, a second request to queue a second media item for playback on the set of devices, where the second user is different than the first
1808
Add the second media item to the playback queue
1810
Initiate playback of the first media item on the set of devices
1812
Initiate playback of the second media item on the set of devices

FIG. 18

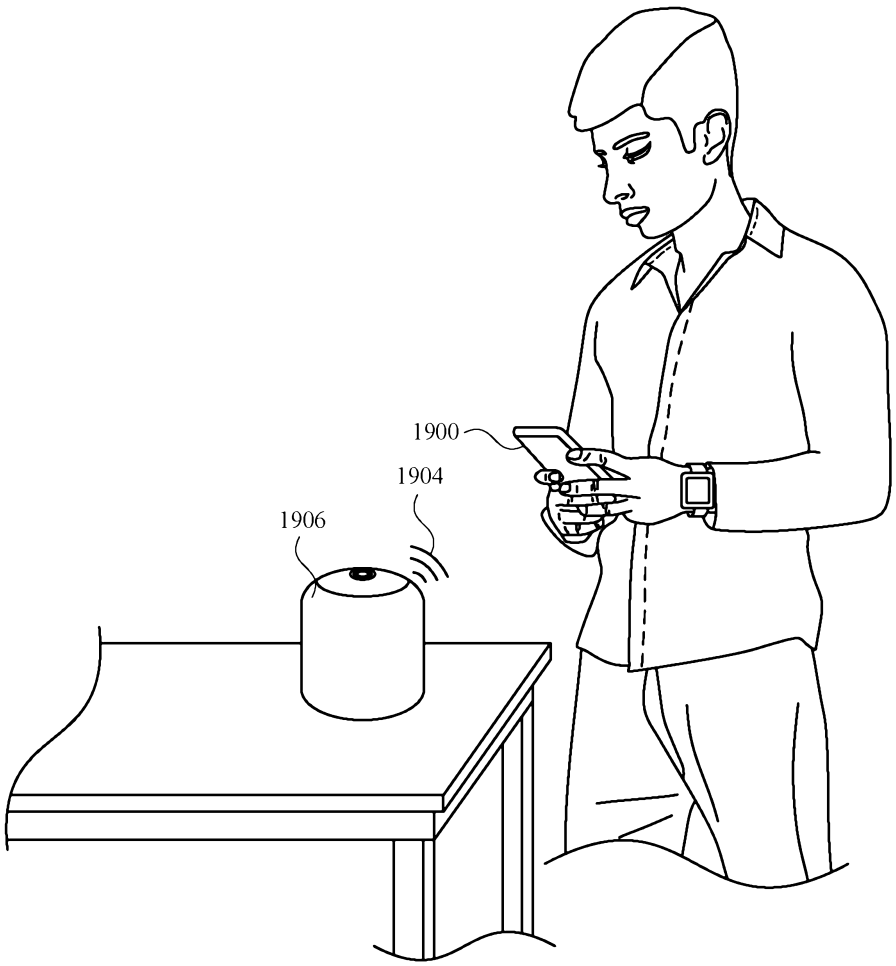


FIG. 19A

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
PHONE
MESSAGES
MAIL
BROWSER

FIG. 19B

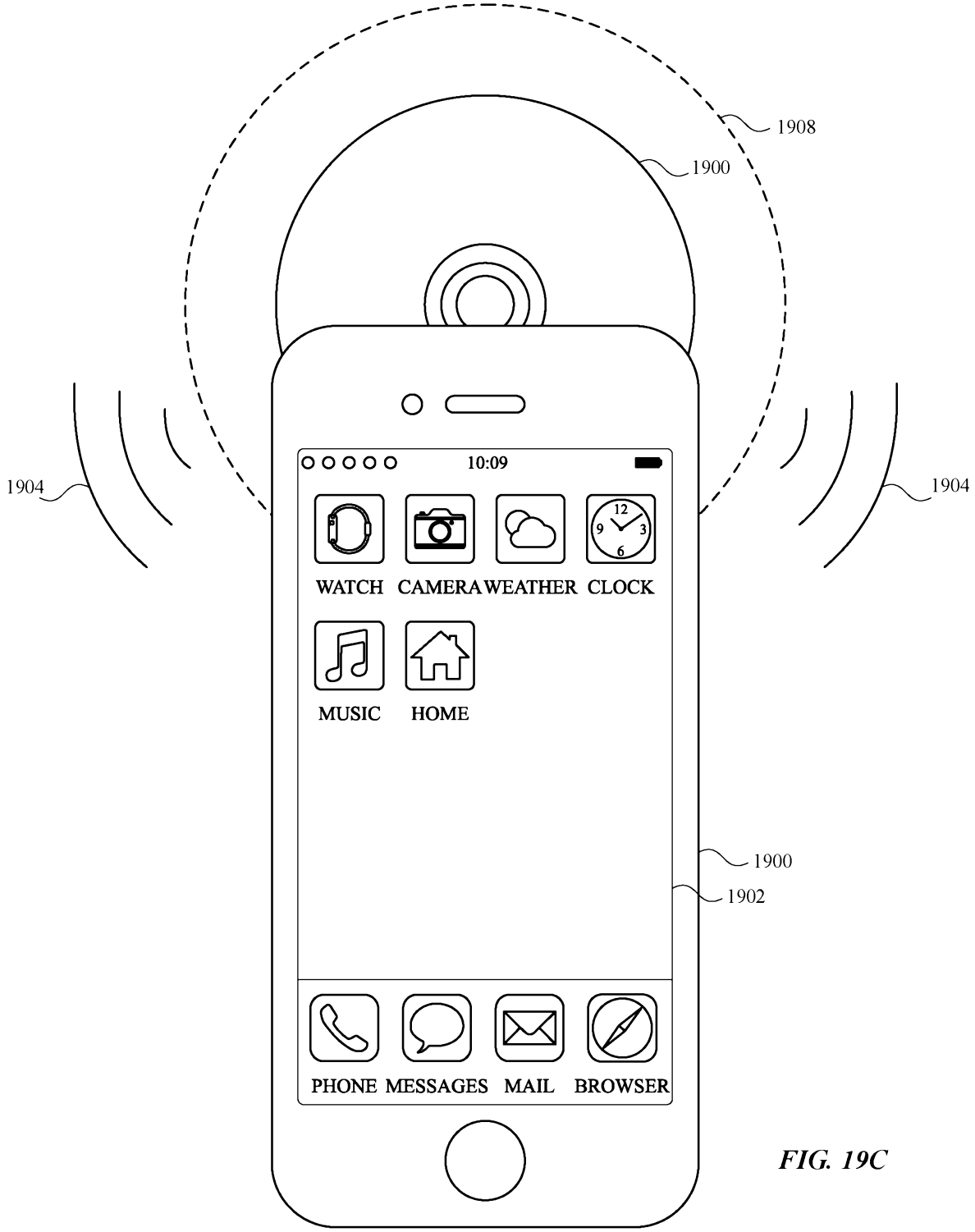

1908
1900
1904
1904
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1900
1902
PHONE
MESSAGES
MAIL
BROWSER

FIG. 19C

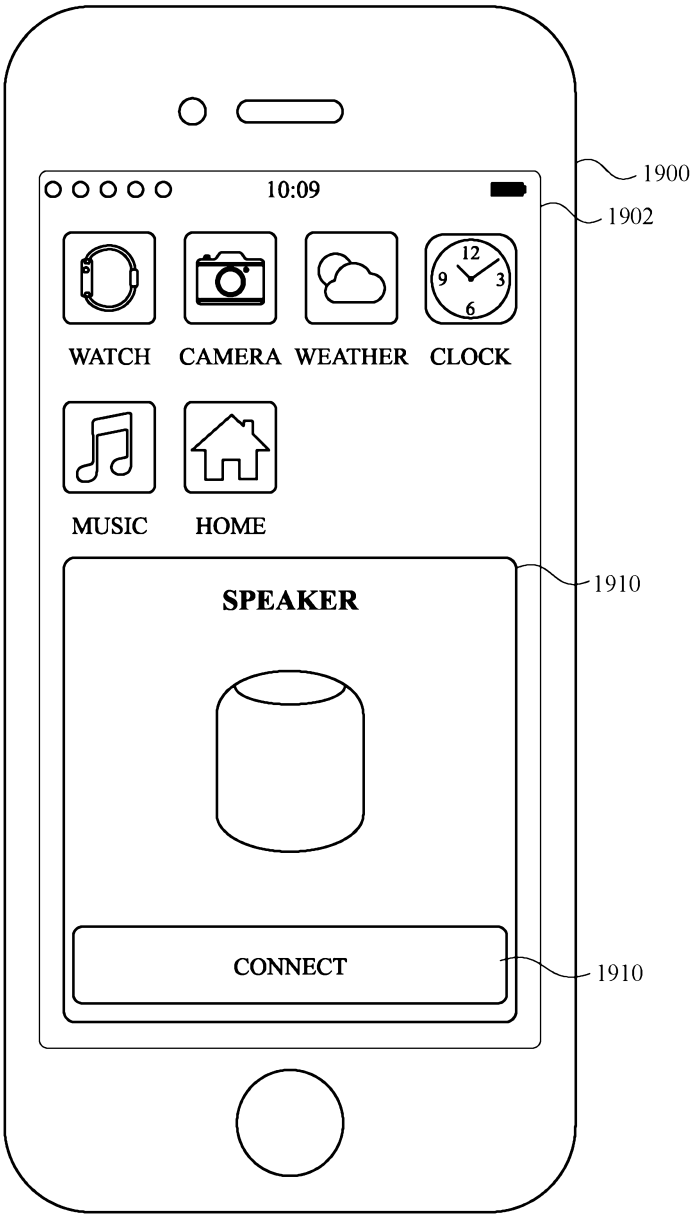

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1910
SPEAKER
CONNECT
1910

FIG. 19D

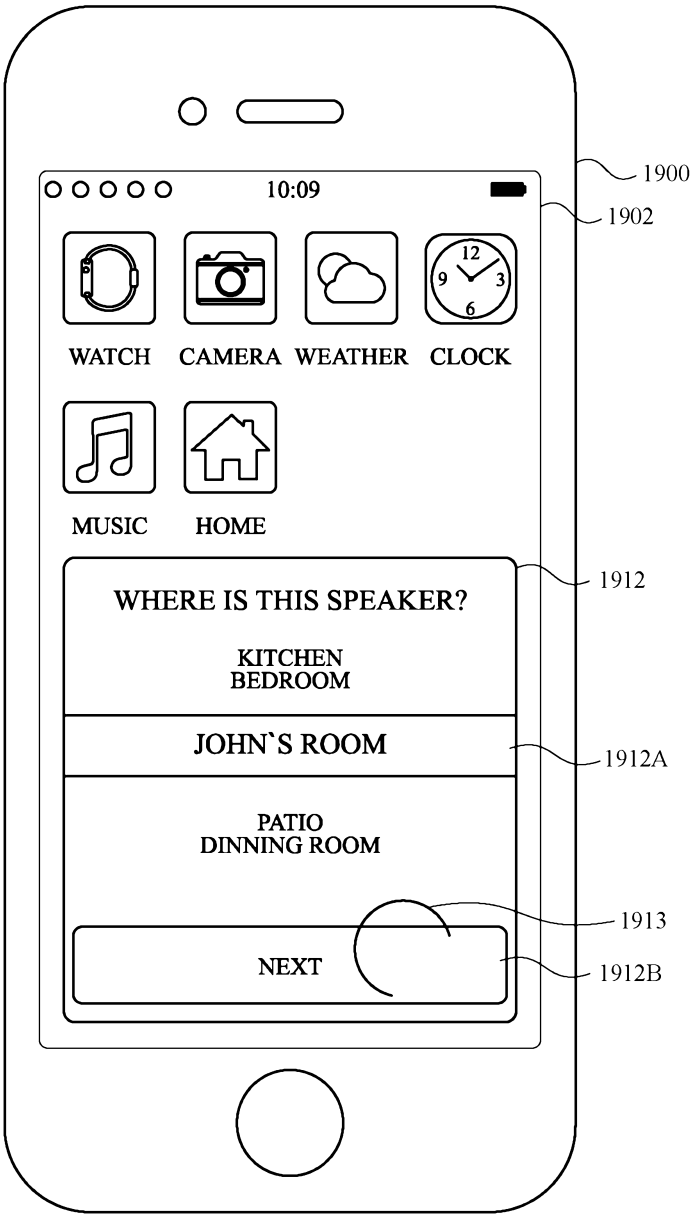

10:09
1900
1902
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
WHERE IS THIS SPEAKER?
KITCHEN
BEDROOM
JOHN`S ROOM
PATIO
DINNING ROOM
NEXT
1912
1912A
1913
1912B

FIG. 19E

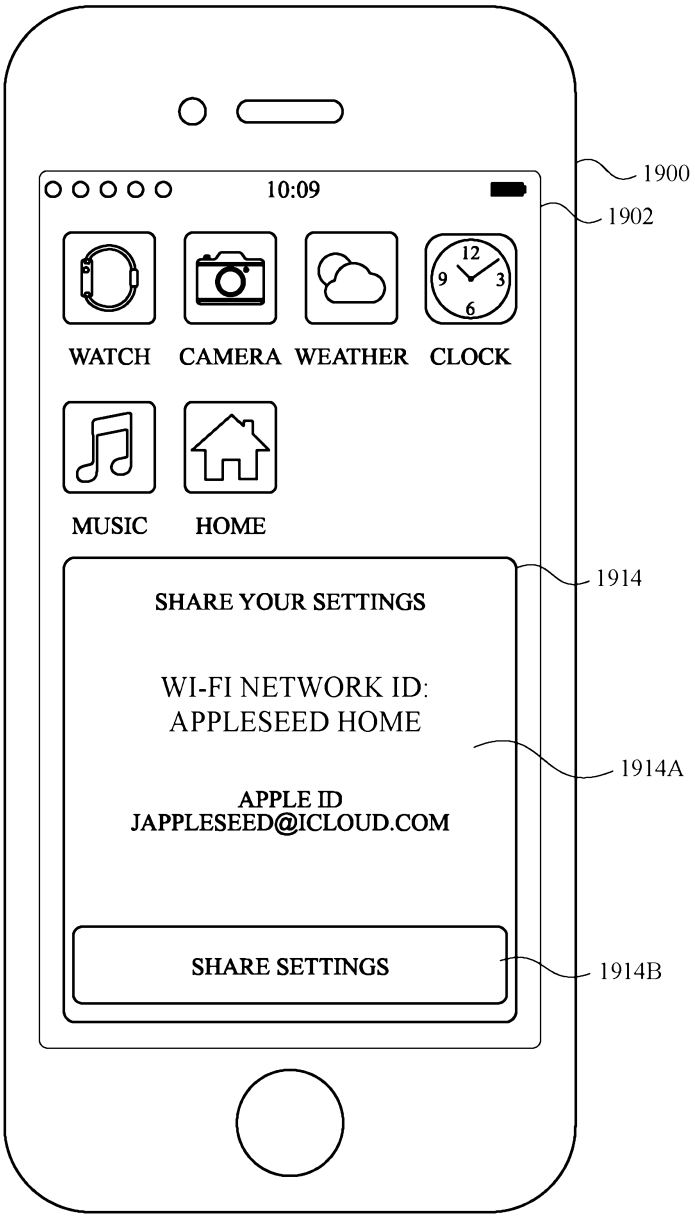

10:09
1900
1902
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
SHARE YOUR SETTINGS
1914
WI-FI NETWORK ID:
APPLESEED HOME
1914A
APPLE ID
JAPPLESEED@ICLOUD.COM
SHARE SETTINGS
1914B

FIG. 19F

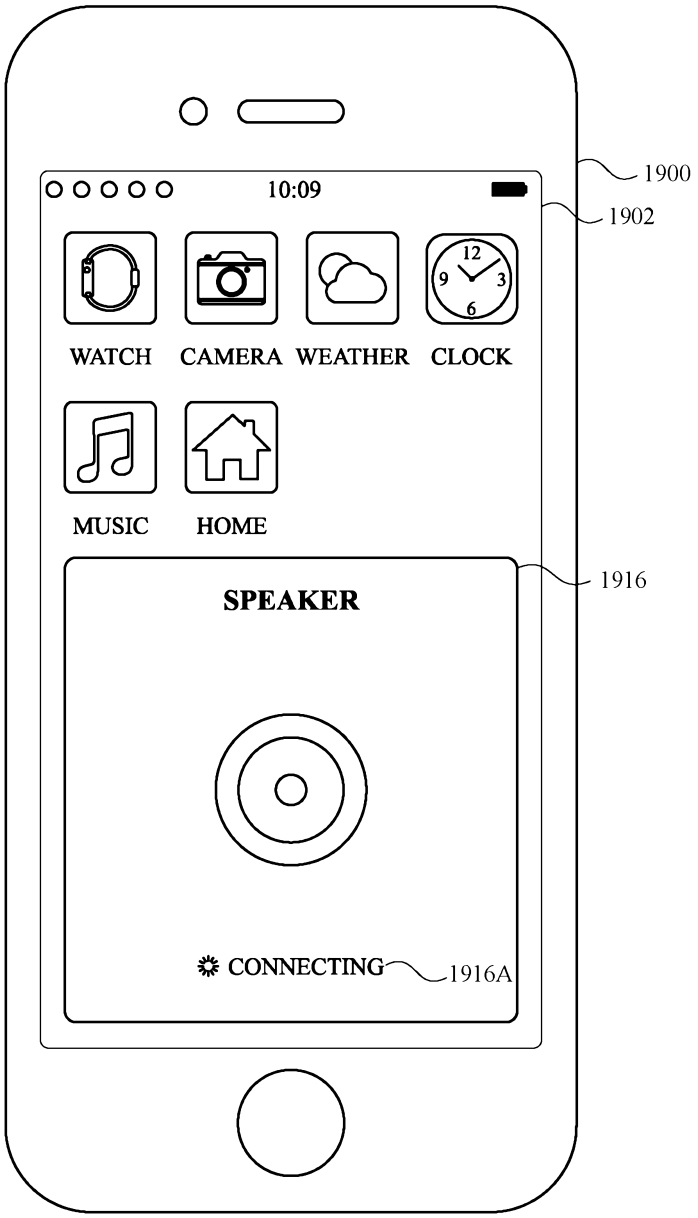

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1916
SPEAKER
CONNECTING
1916A

FIG. 19G

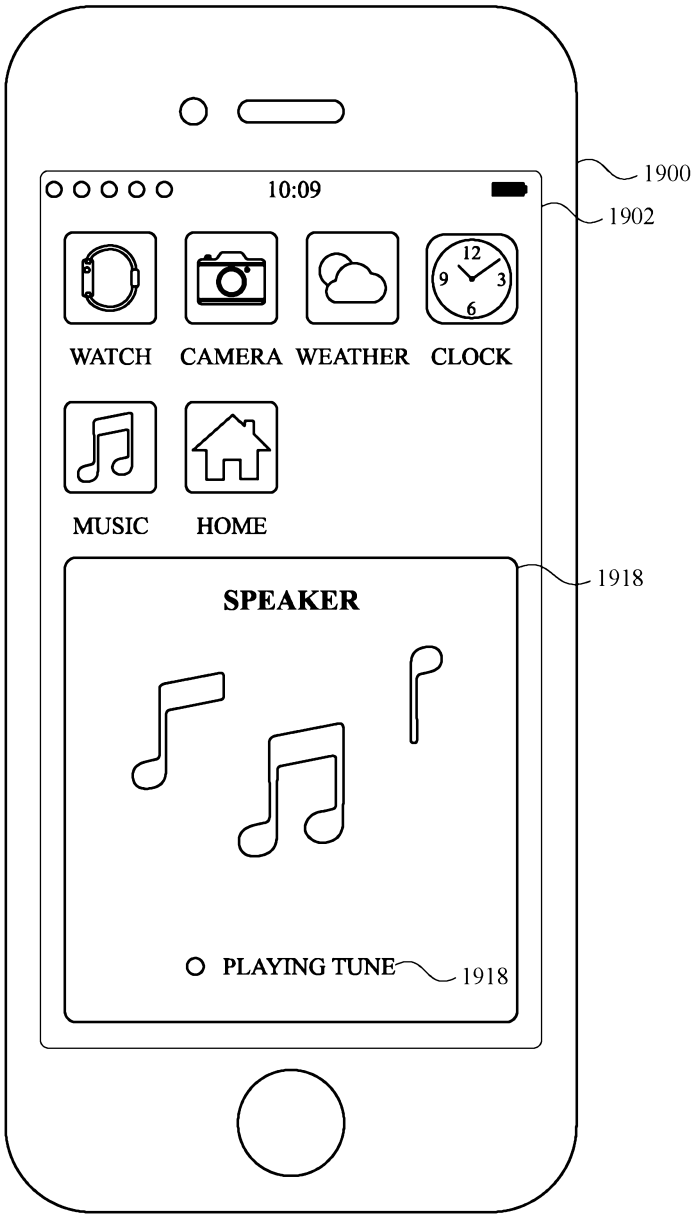

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1918
SPEAKER
PLAYING TUNE
1918

FIG. 19H

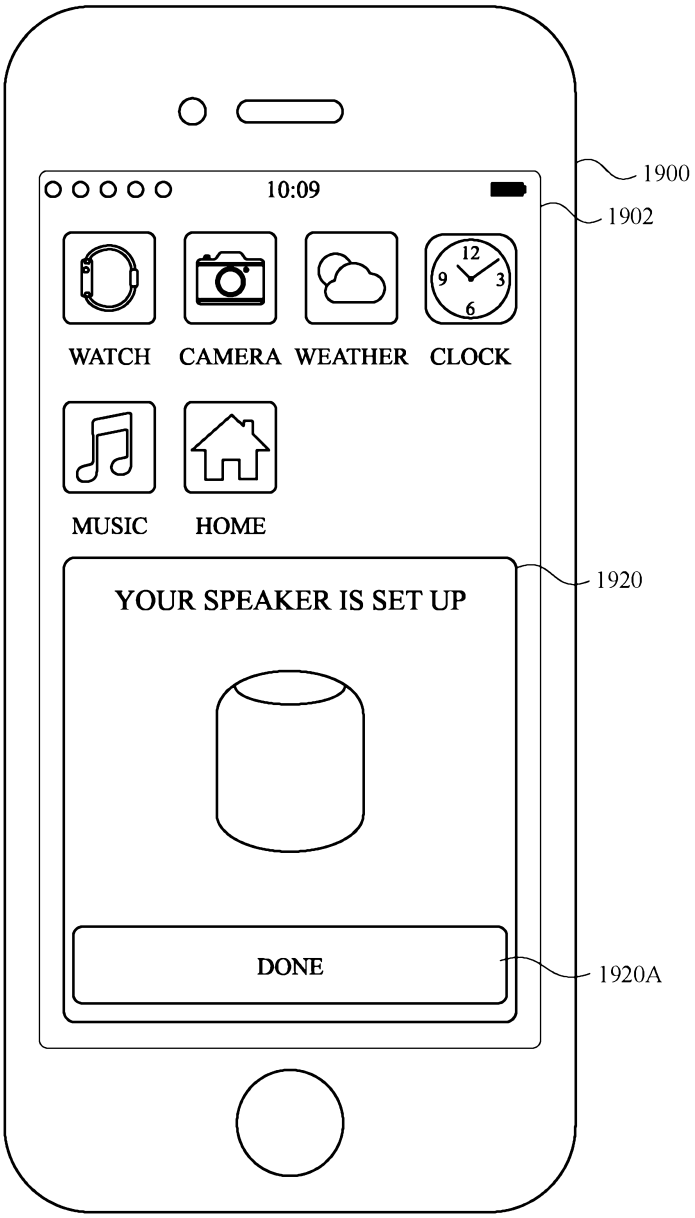

1900
10:09
1902
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1920
YOUR SPEAKER IS SET UP
DONE
1920A

FIG. 19I

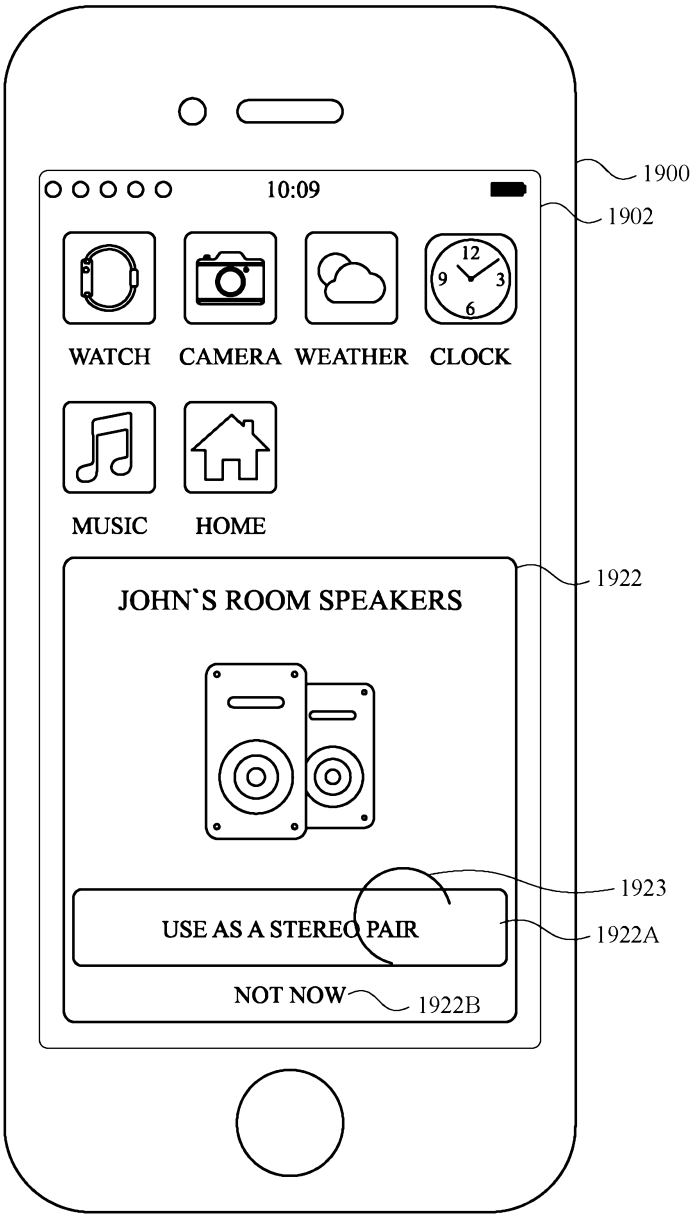

10:09
1900
1902
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
JOHN`S ROOM SPEAKERS
1922
1923
USE AS A STEREO PAIR
1922A
NOT NOW
1922B

FIG. 19J

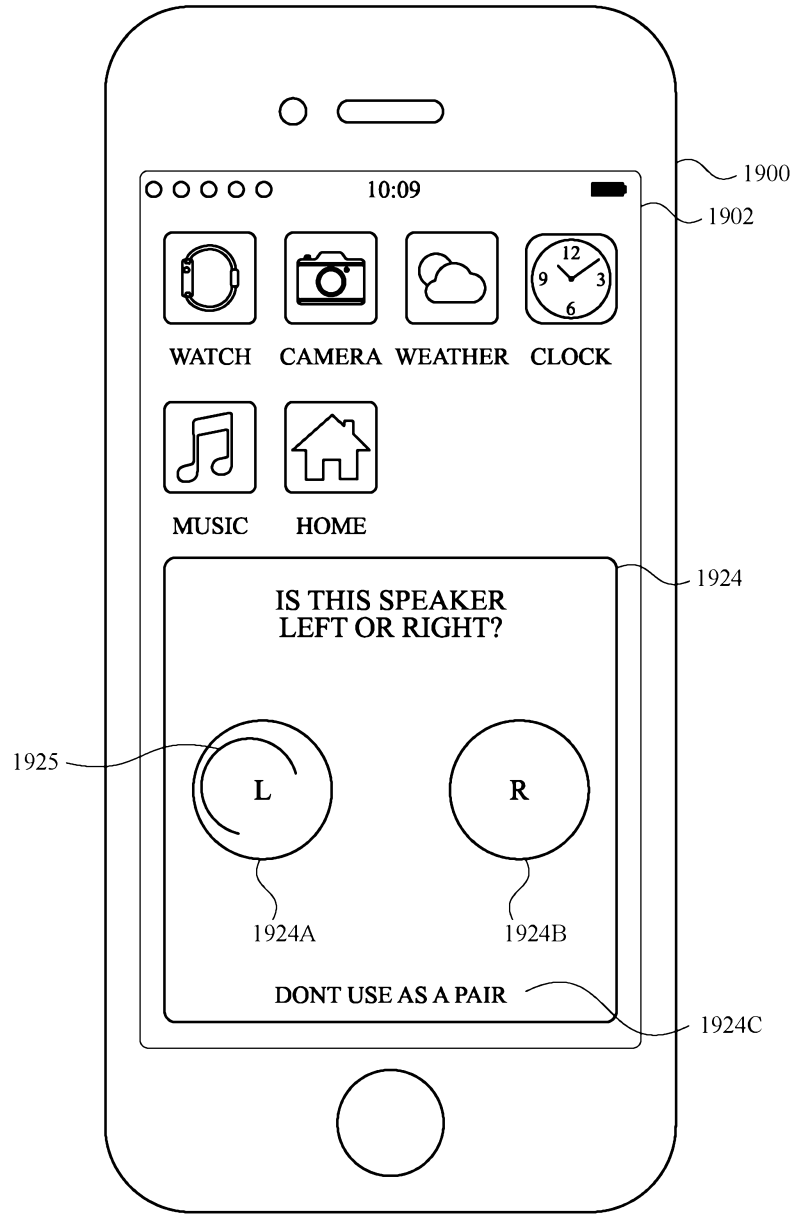

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1924
IS THIS SPEAKER
LEFT OR RIGHT?
1925
L
R
1924A
1924B
DONT USE AS A PAIR
1924C

FIG. 19K

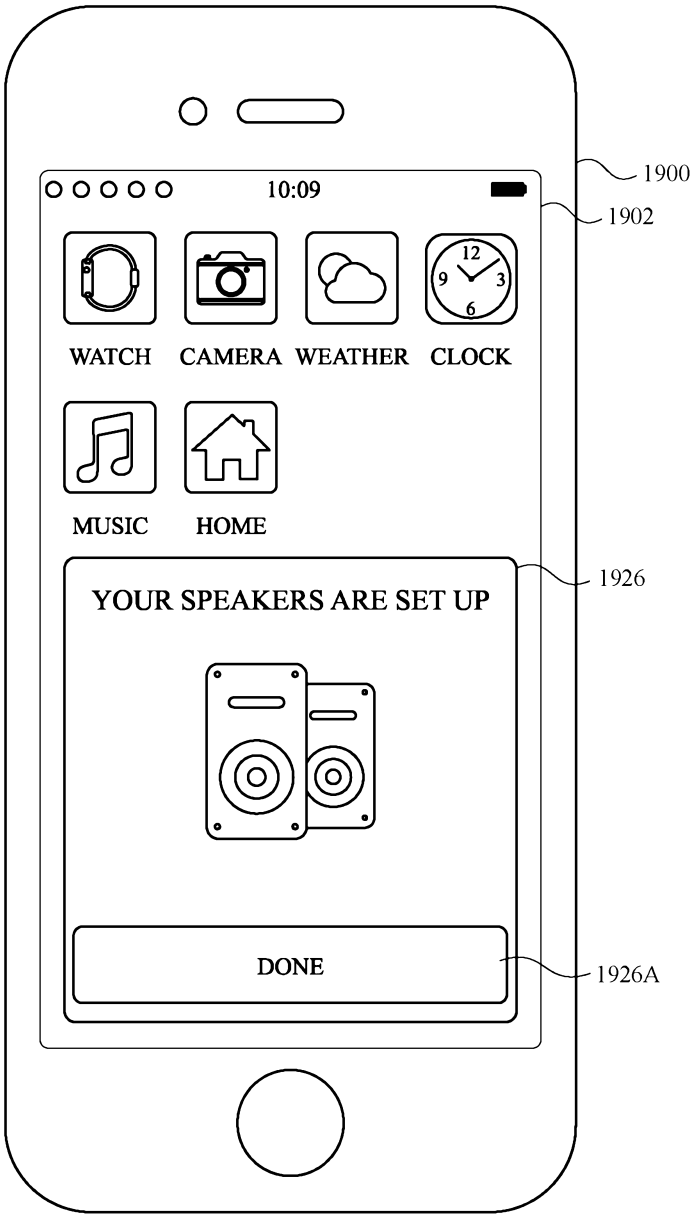

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1926
YOUR SPEAKERS ARE SET UP
DONE
1926A

FIG. 19L

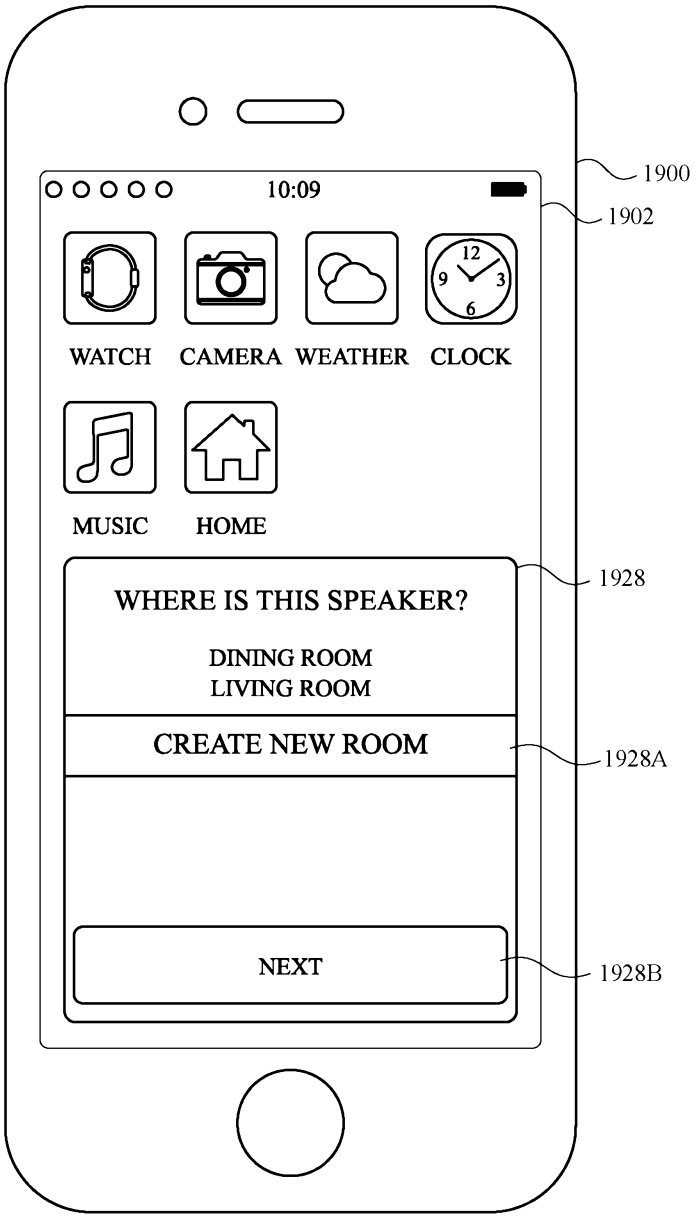

1900
10:09
1902
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1928
WHERE IS THIS SPEAKER?
DINING ROOM
LIVING ROOM
CREATE NEW ROOM
1928A
NEXT
1928B

FIG. 19M

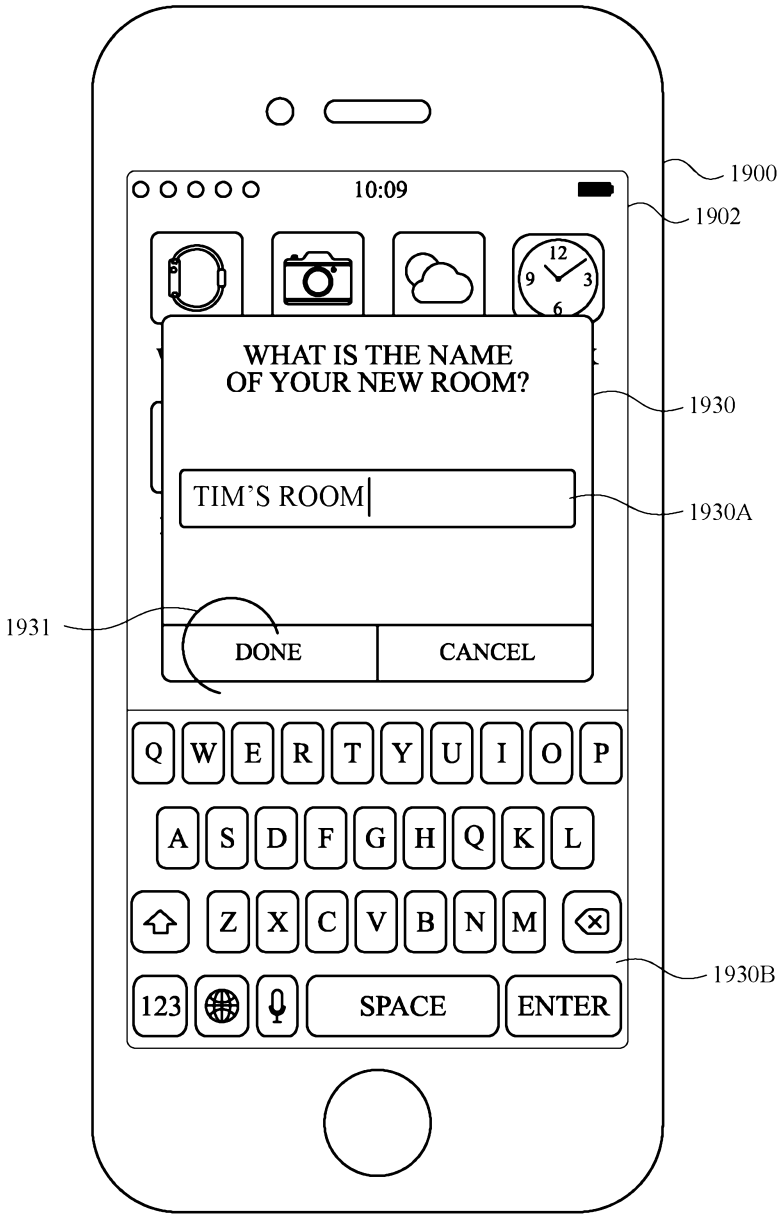

10:09
1900
1902
WHAT IS THE NAME
OF YOUR NEW ROOM?
1930
TIM'S ROOM
1930A
1931
DONE
CANCEL
Q W E R T Y U I O P
A S D F G H Q K L
Z X C V B N M
1930B
123
SPACE
ENTER

FIG. 19N

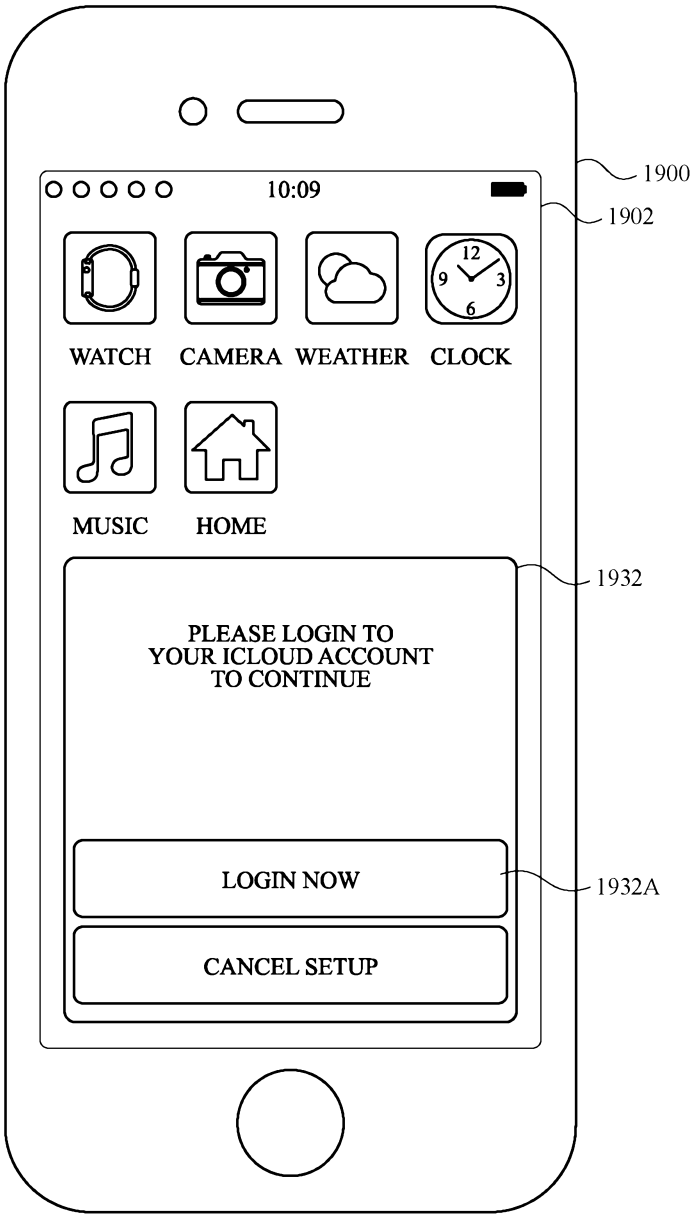

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1932
PLEASE LOGIN TO
YOUR ICLOUD ACCOUNT
TO CONTINUE
LOGIN NOW
1932A
CANCEL SETUP

FIG. 19O

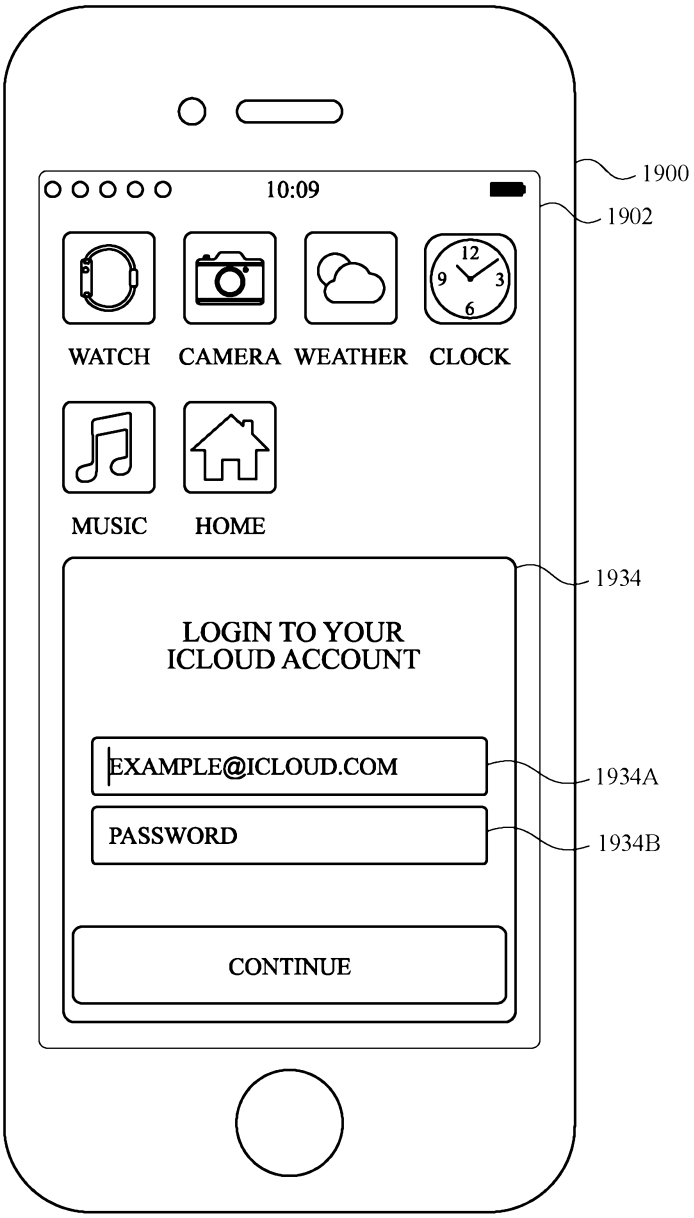

1900
1902
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
1934
LOGIN TO YOUR
ICLOUD ACCOUNT
EXAMPLE@ICLOUD.COM
1934A
PASSWORD
1934B
CONTINUE

FIG. 19P

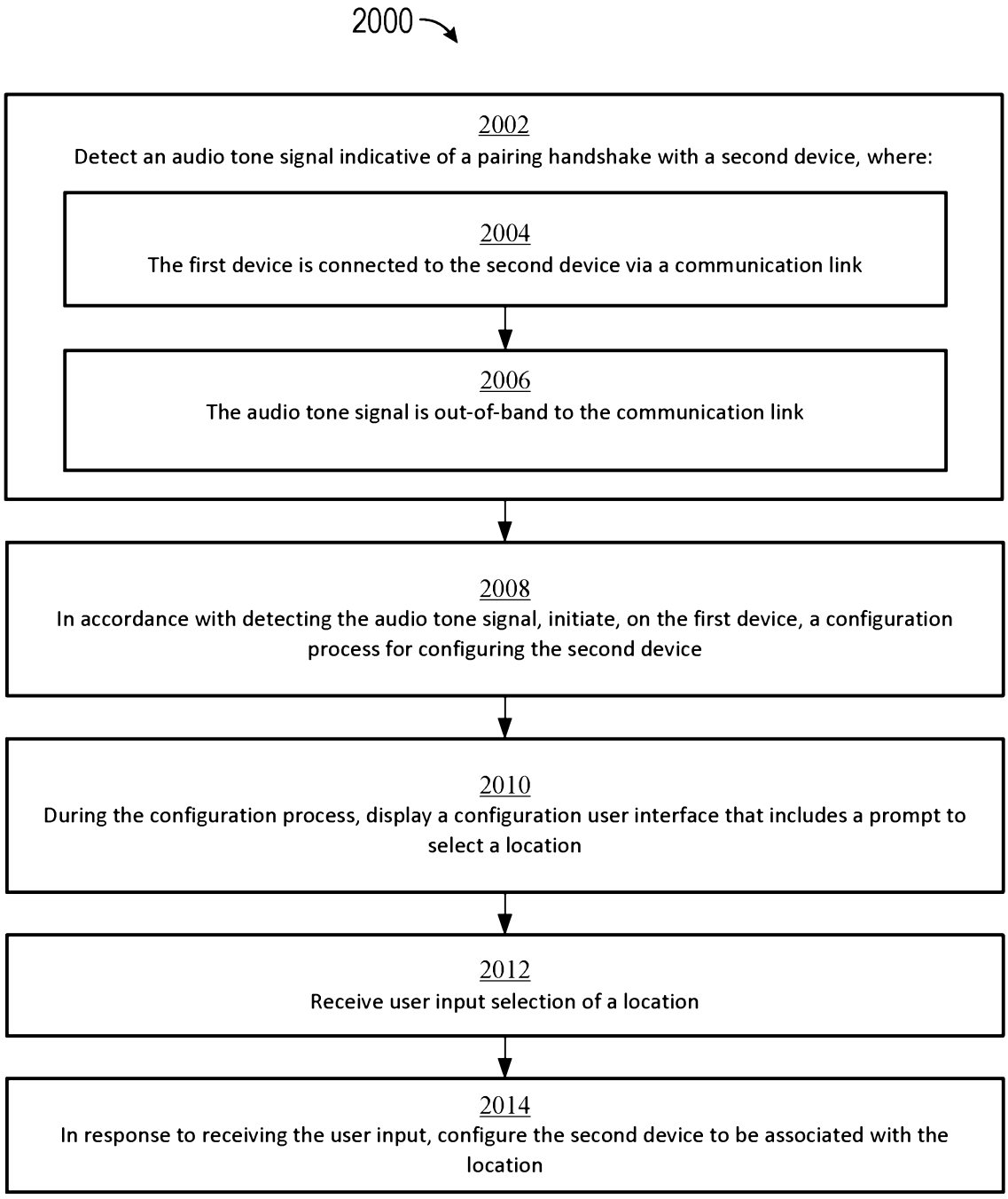

2000
2002
Detect an audio tone signal indicative of a pairing handshake with a second device, where:
2004
The first device is connected to the second device via a communication link
2006
The audio tone signal is out-of-band to the communication link
2008
In accordance with detecting the audio tone signal, initiate, on the first device, a configuration process for configuring the second device
2010
During the configuration process, display a configuration user interface that includes a prompt to select a location
2012
Receive user input selection of a location
2014
In response to receiving the user input, configure the second device to be associated with the location

FIG. 20

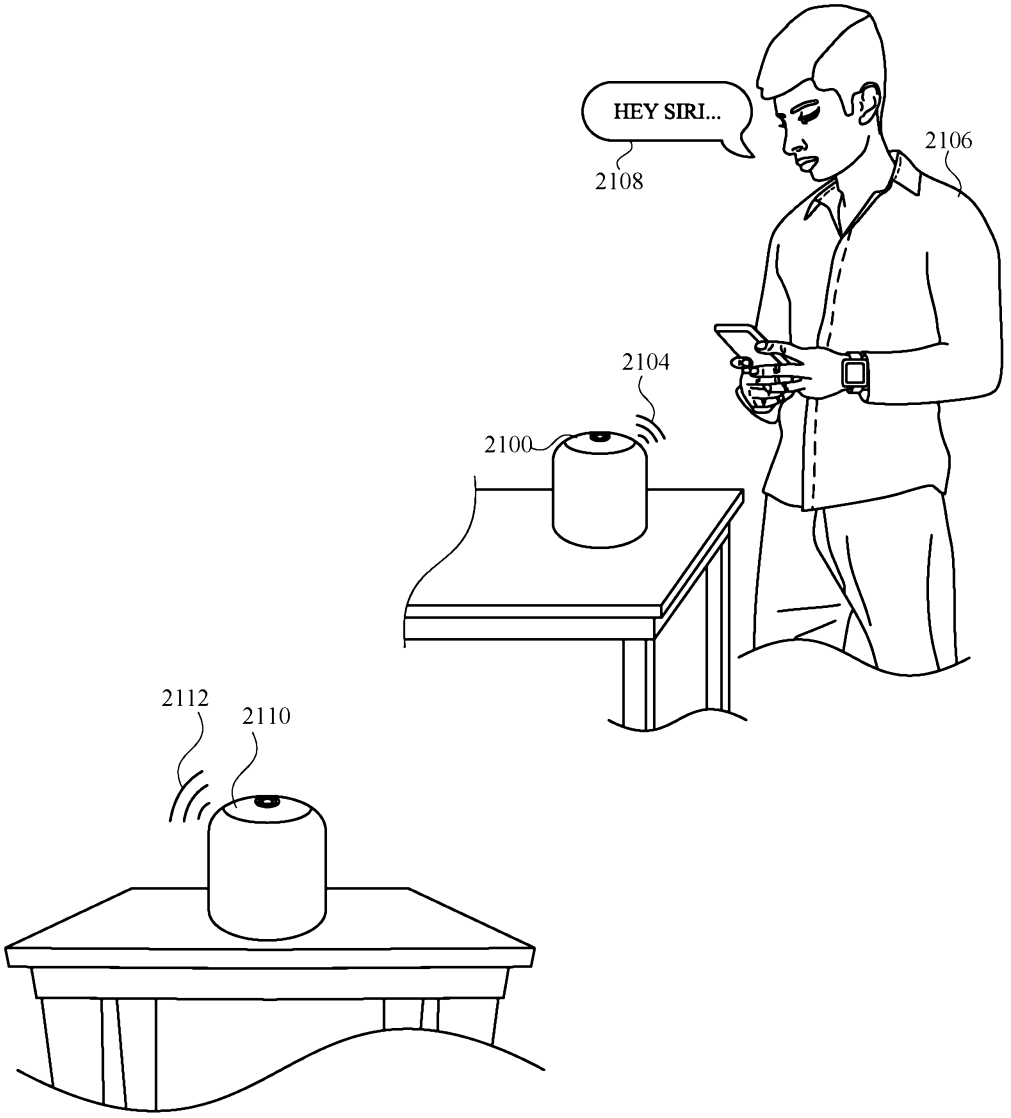


FIG. 21A

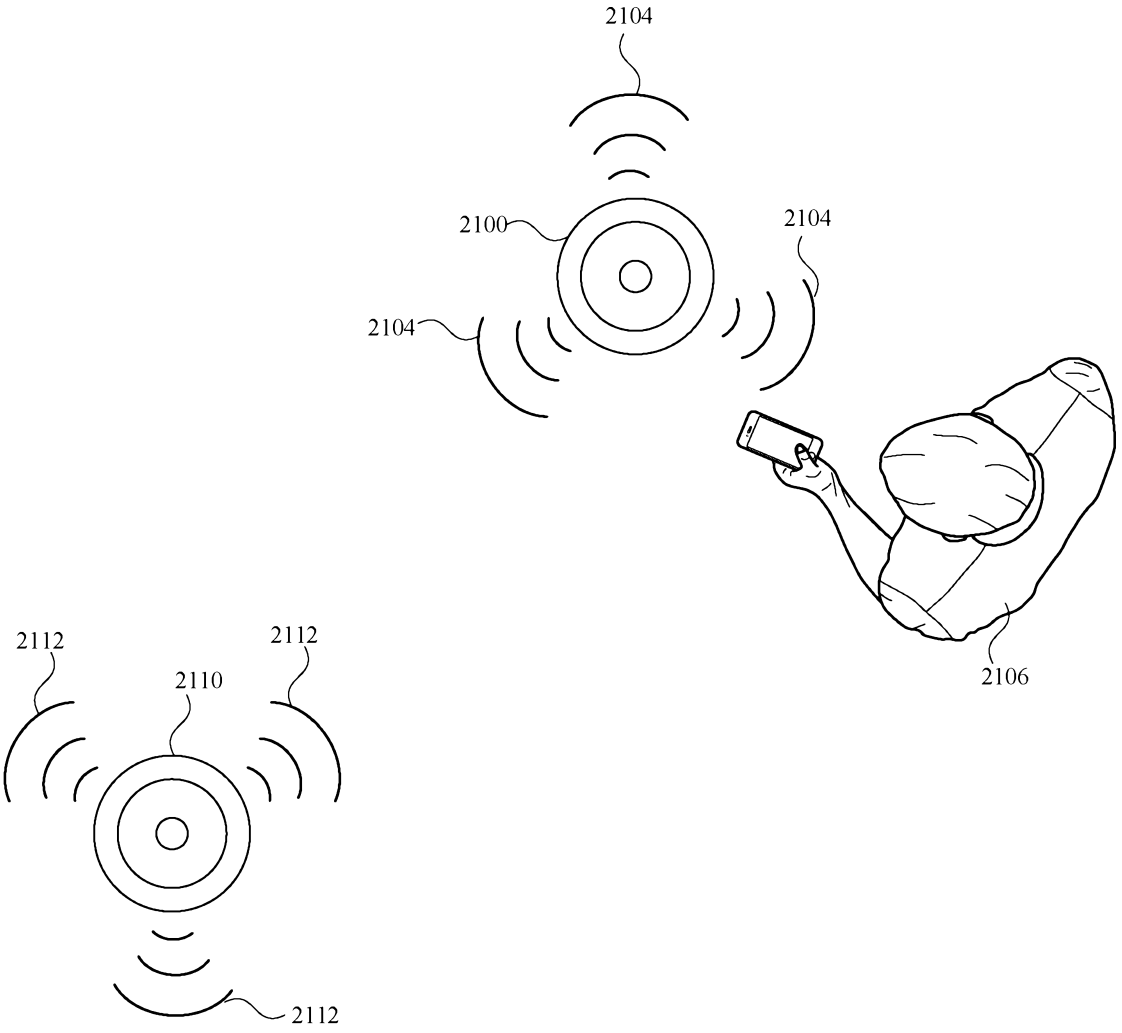


FIG. 21B

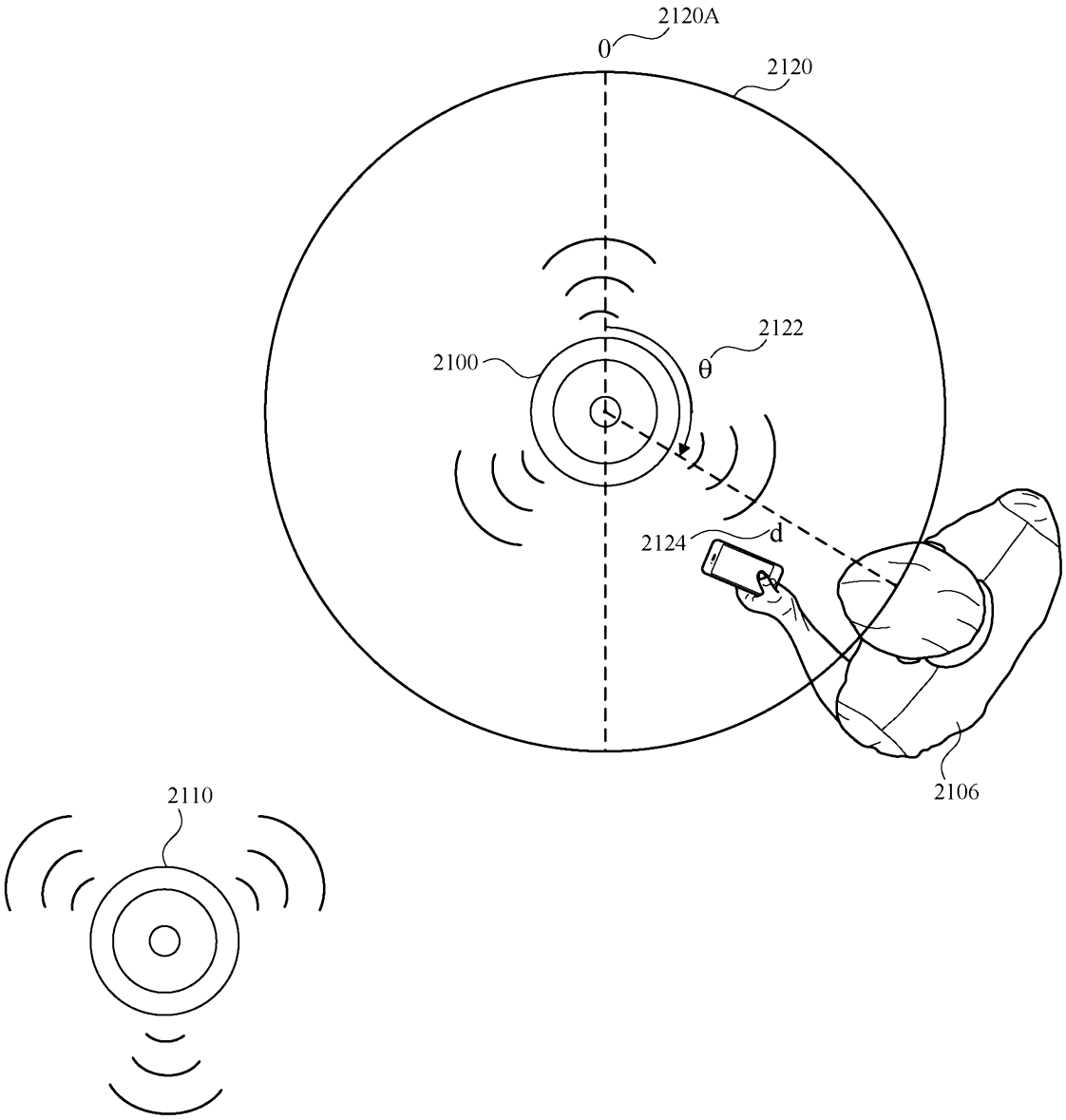


FIG. 21C

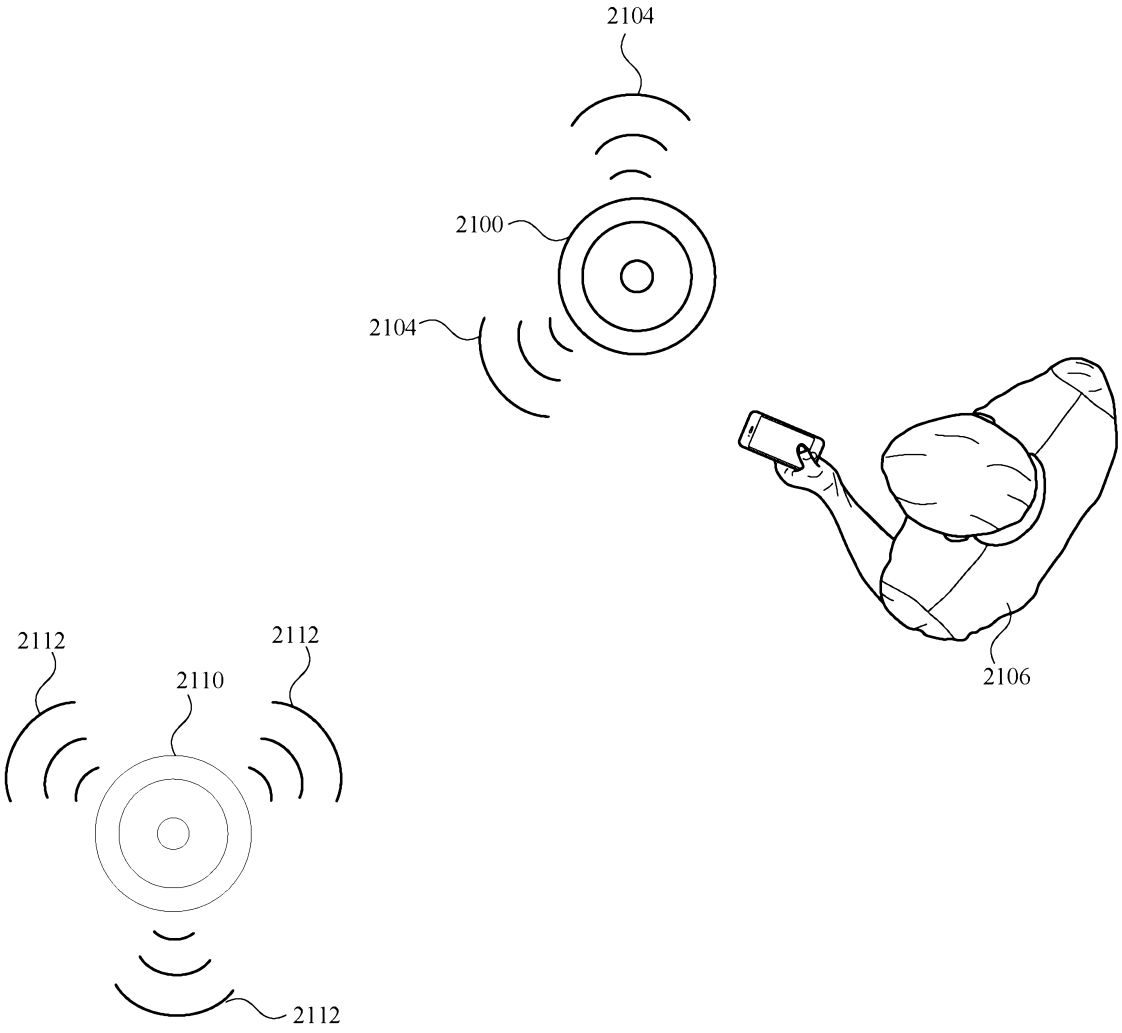


FIG. 21D

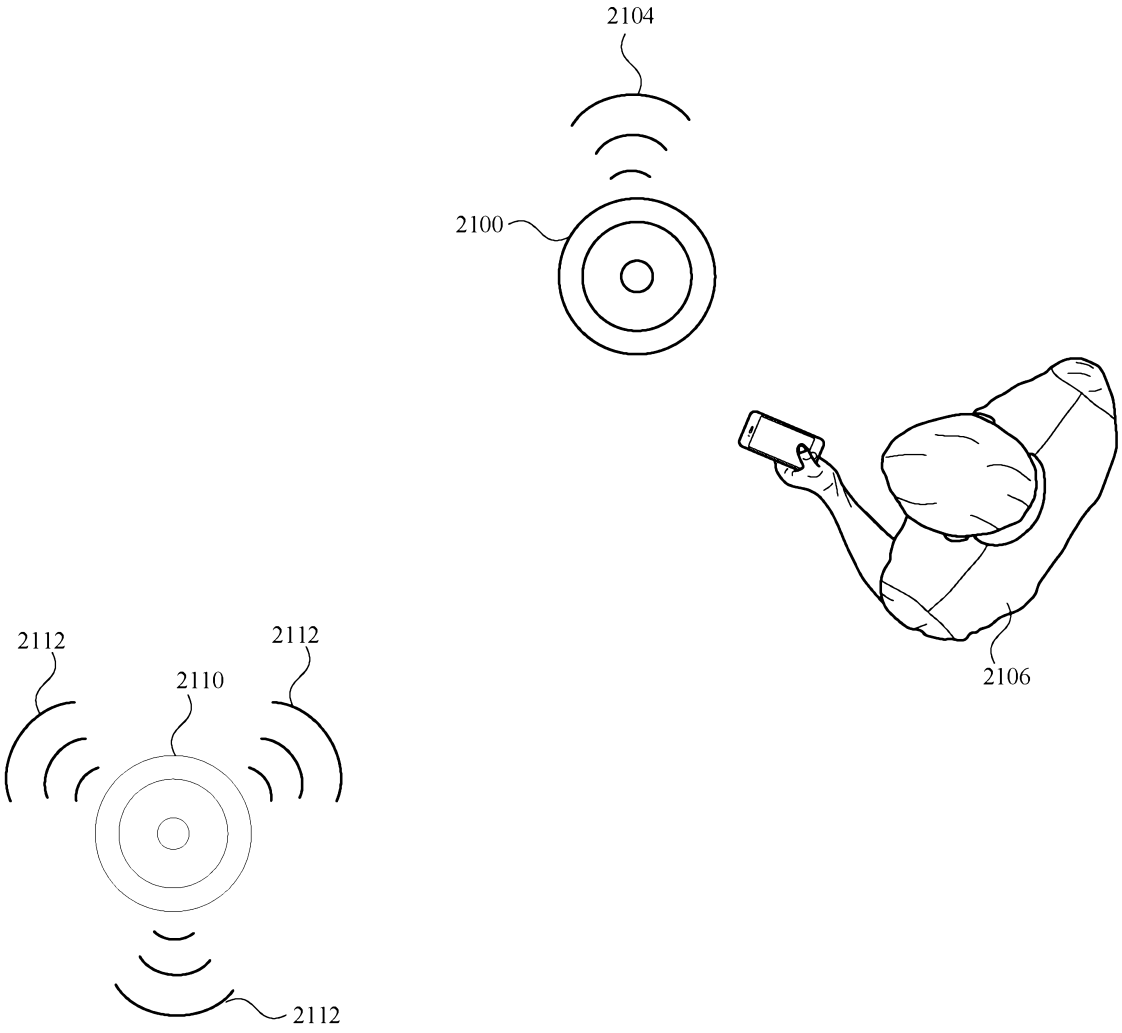


FIG. 21E

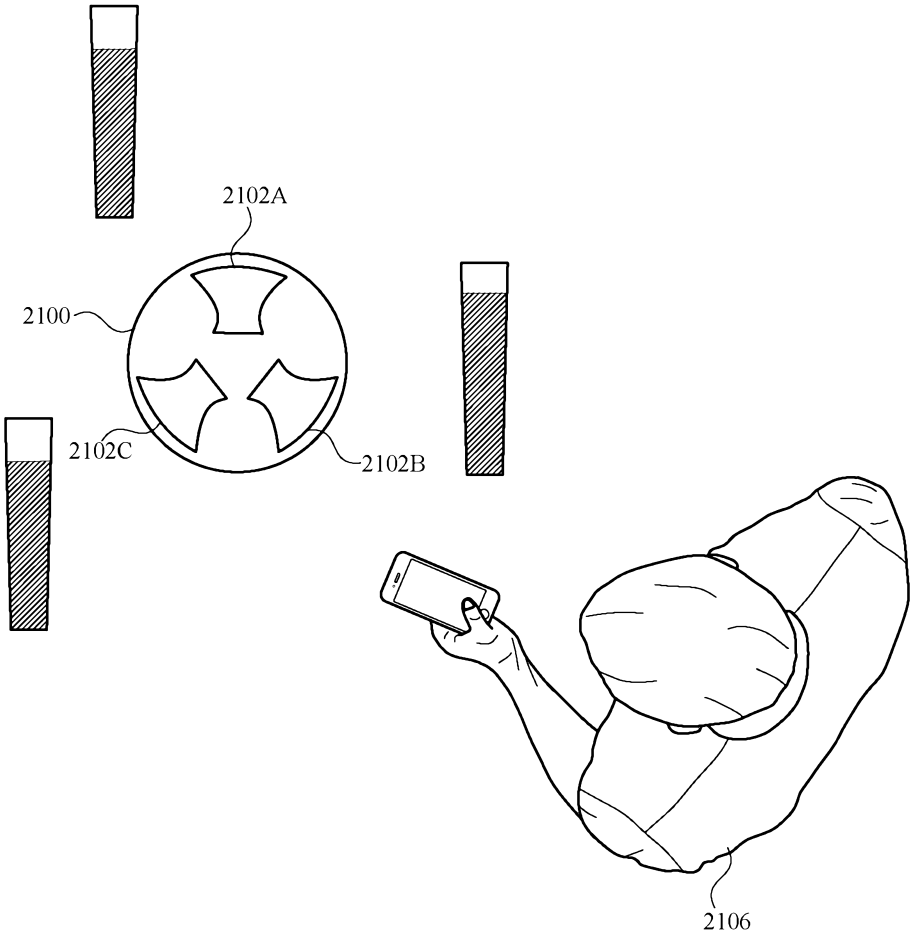


FIG. 21F

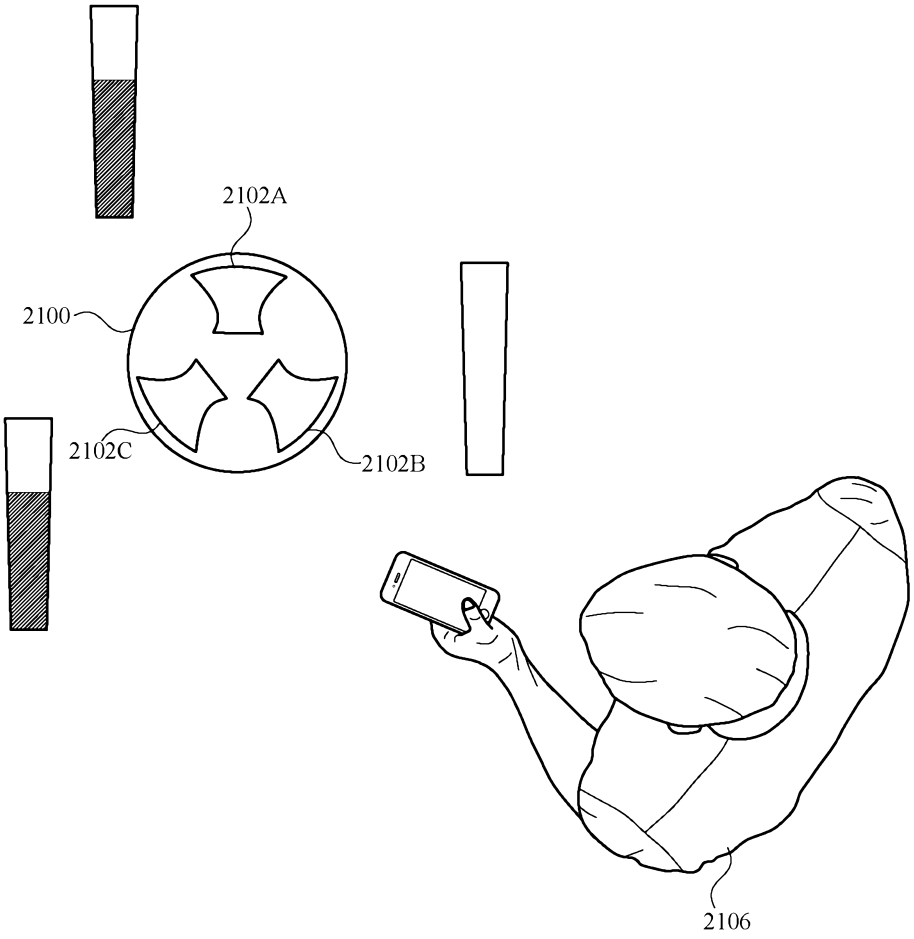


FIG. 21G

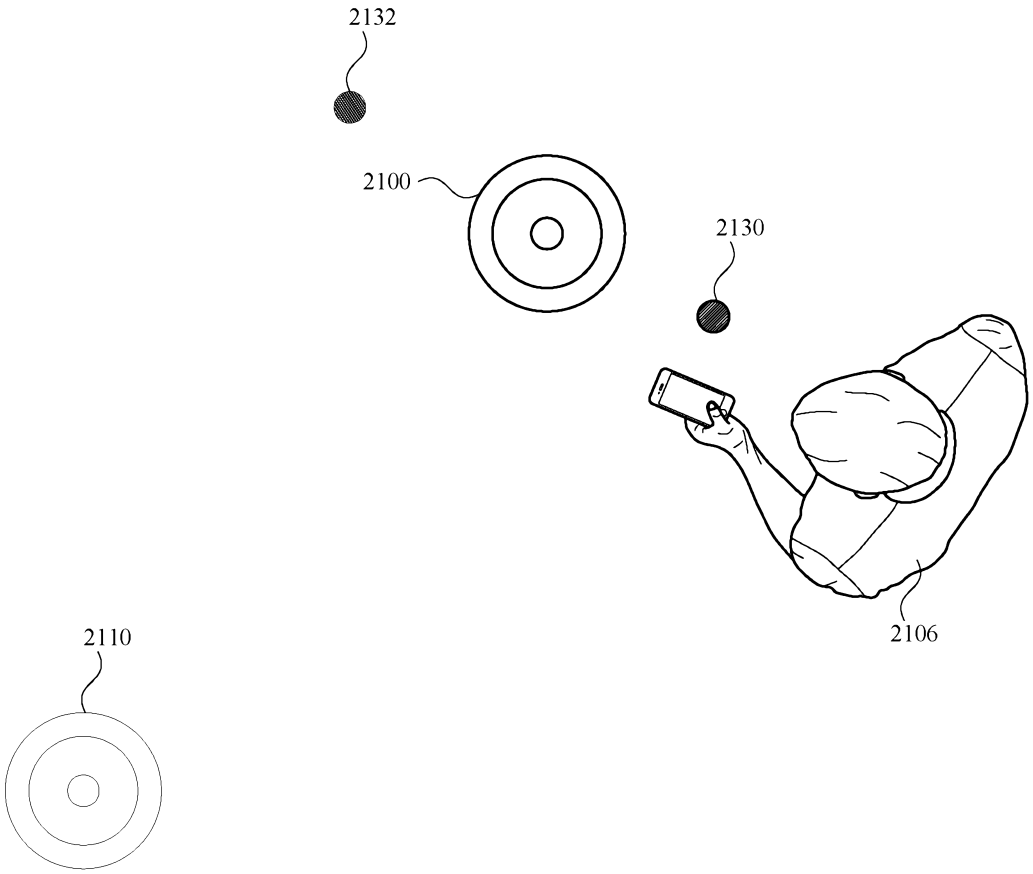


FIG. 21H

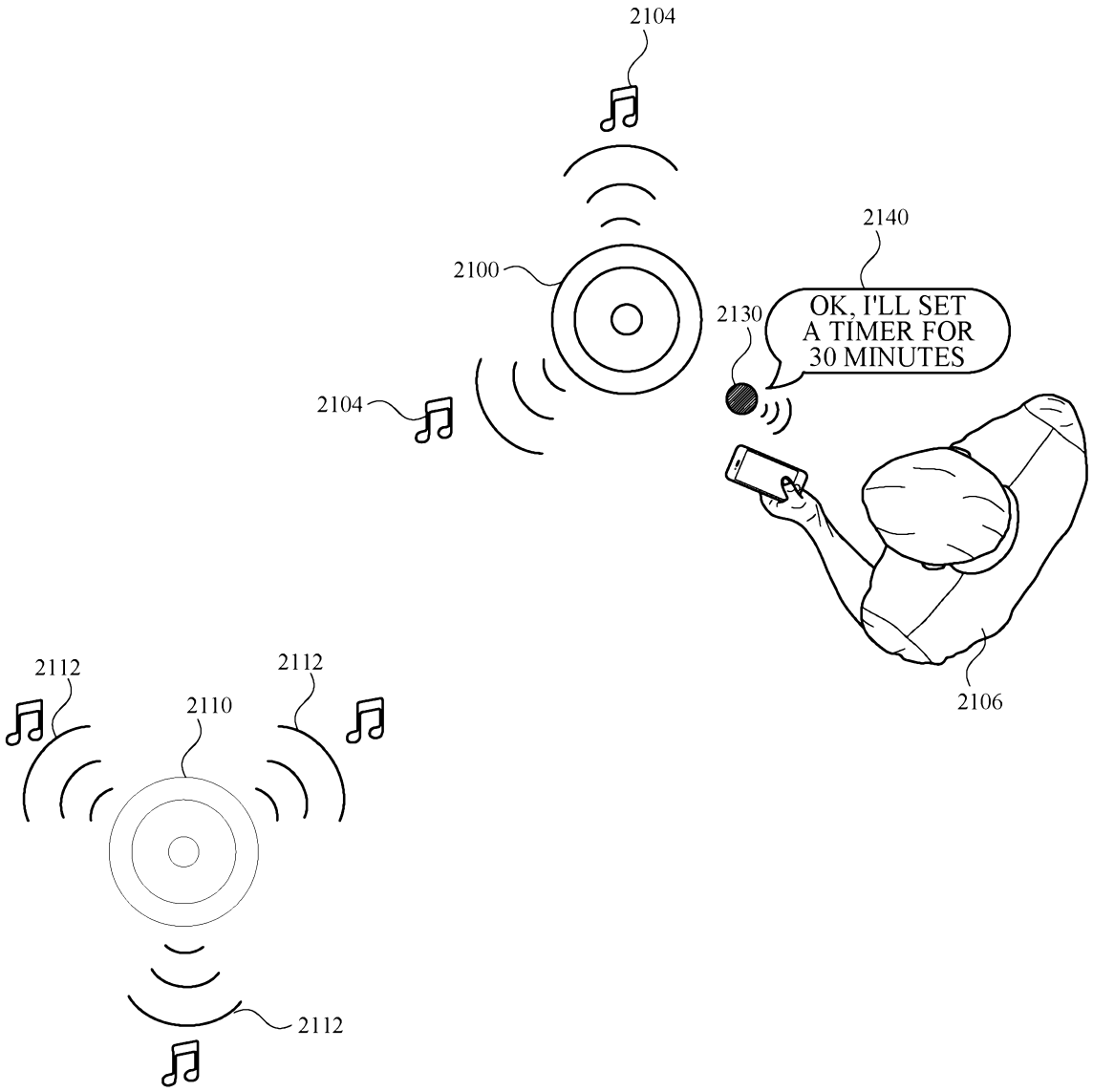

2104
2140
2100
OK, I'LL SET
A TIMER FOR
30 MINUTES
2130
2104
2112
2112
2110
2106
2112

FIG. 21I

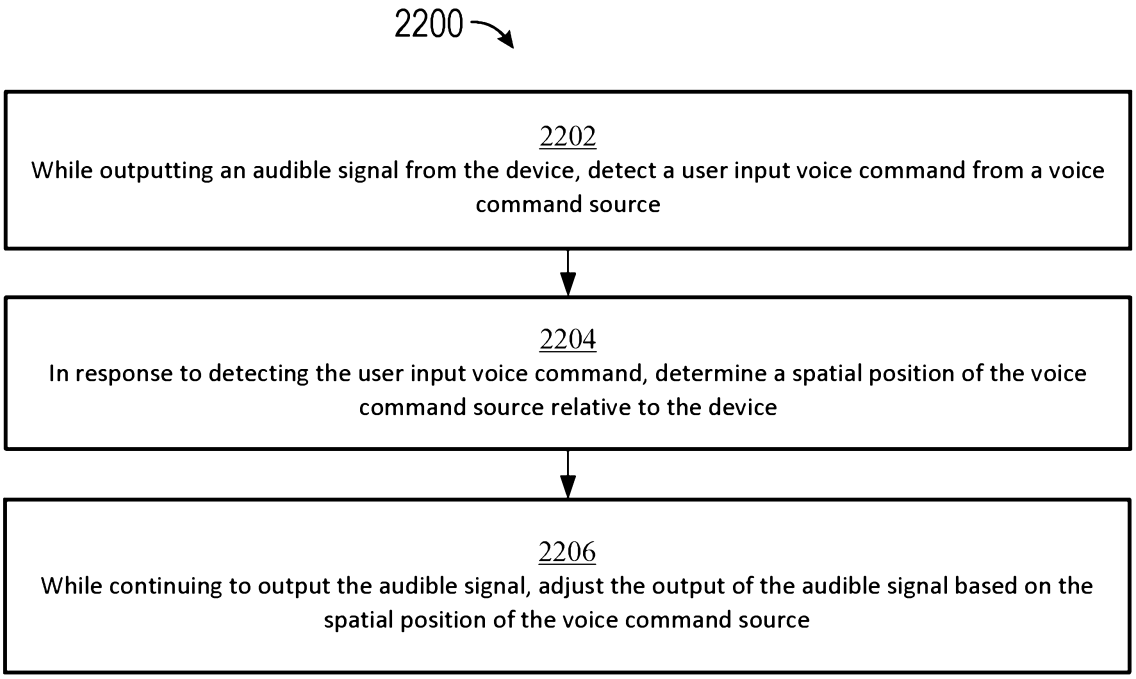

2200
2202
While outputting an audible signal from the device, detect a user input voice command from a voice command source
2204
In response to detecting the user input voice command, determine a spatial position of the voice command source relative to the device
2206
While continuing to output the audible signal, adjust the output of the audible signal based on the spatial position of the voice command source

FIG. 22

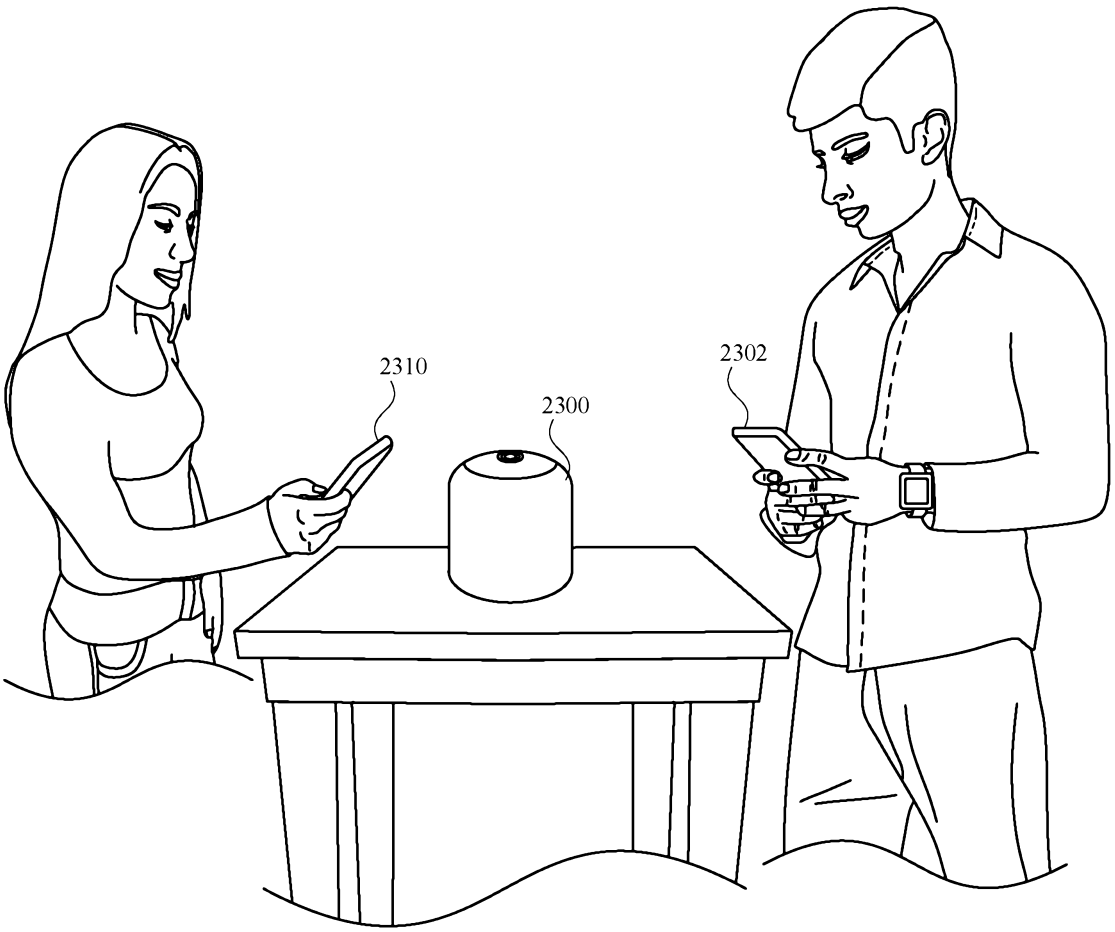


FIG. 23A

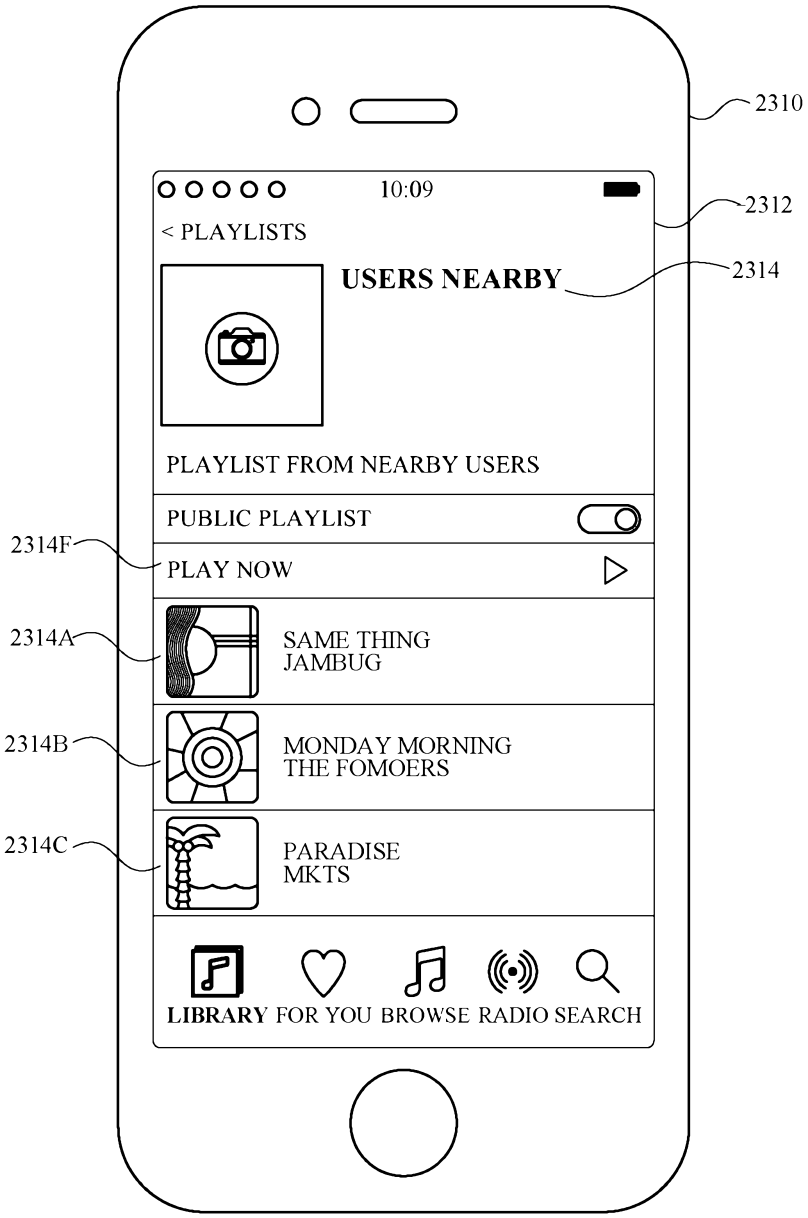

2310
10:09
2312
< PLAYLISTS
USERS NEARBY
2314
PLAYLIST FROM NEARBY USERS
PUBLIC PLAYLIST
2314F
PLAY NOW
2314A
SAME THING
JAMBUG
2314B
MONDAY MORNING
THE FOMOERS
2314C
PARADISE
MKTS
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 23B

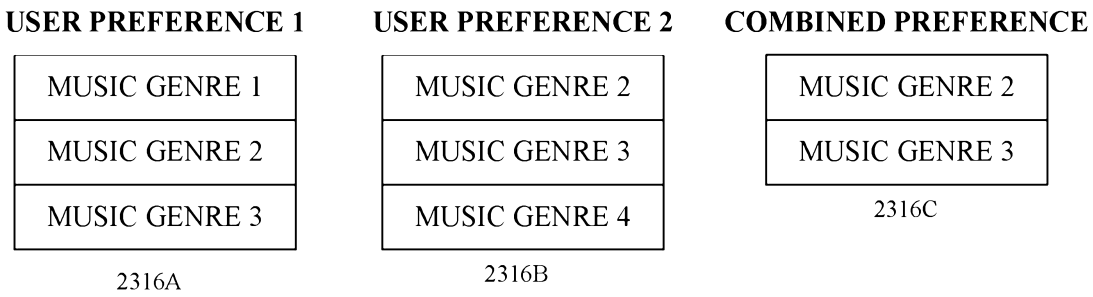

USER PREFERENCE 1
MUSIC GENRE 1
MUSIC GENRE 2
MUSIC GENRE 3
2316A
USER PREFERENCE 2
MUSIC GENRE 2
MUSIC GENRE 3
MUSIC GENRE 4
2316B
COMBINED PREFERENCE
MUSIC GENRE 2
MUSIC GENRE 3
2316C

FIG. 23C

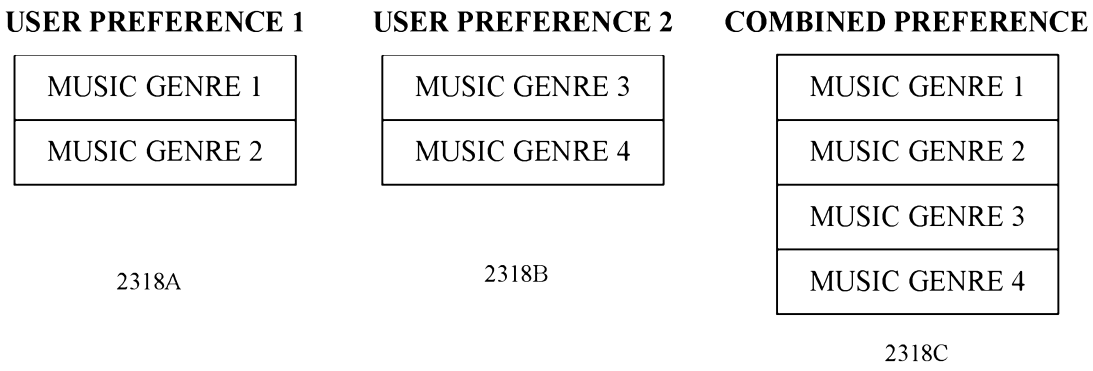

USER PREFERENCE 1
MUSIC GENRE 1
MUSIC GENRE 2
2318A
USER PREFERENCE 2
MUSIC GENRE 3
MUSIC GENRE 4
2318B
COMBINED PREFERENCE
MUSIC GENRE 1
MUSIC GENRE 2
MUSIC GENRE 3
MUSIC GENRE 4
2318C

FIG. 23D

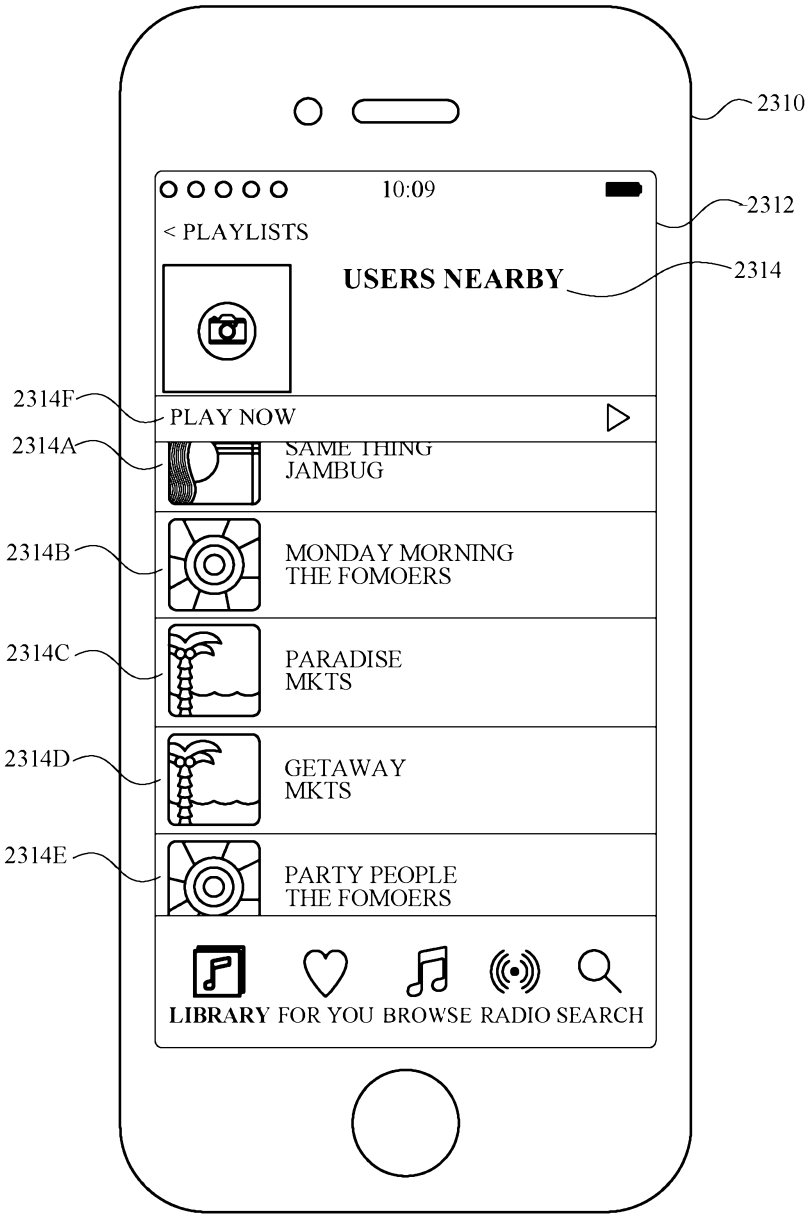

2310
10:09
2312
< PLAYLISTS
USERS NEARBY
2314
2314F
PLAY NOW
2314A
SAME THING
JAMBUG
2314B
MONDAY MORNING
THE FOMOERS
2314C
PARADISE
MKTS
2314D
GETAWAY
MKTS
2314E
PARTY PEOPLE
THE FOMOERS
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 23E

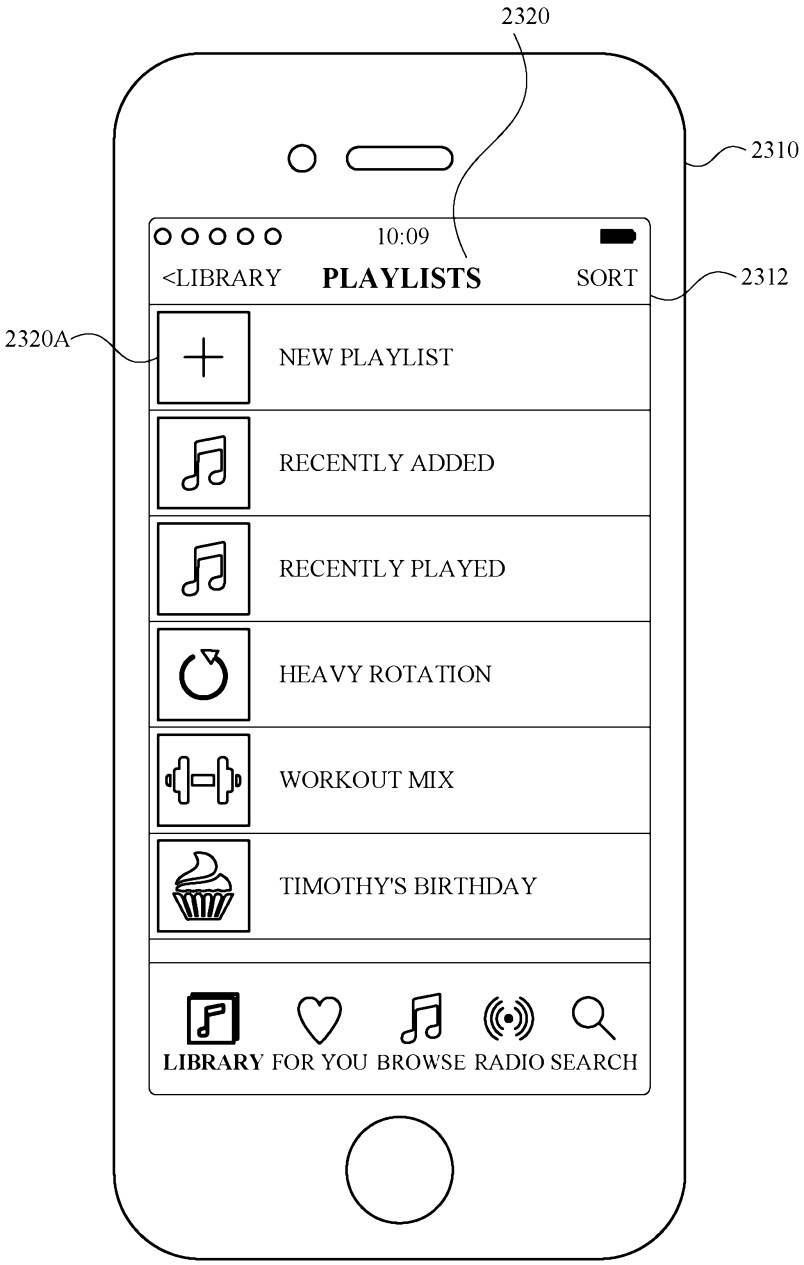

2320
2310
10:09
<LIBRARY
PLAYLISTS
SORT
2312
2320A
NEW PLAYLIST
RECENTLY ADDED
RECENTLY PLAYED
HEAVY ROTATION
WORKOUT MIX
TIMOTHY'S BIRTHDAY
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 23F

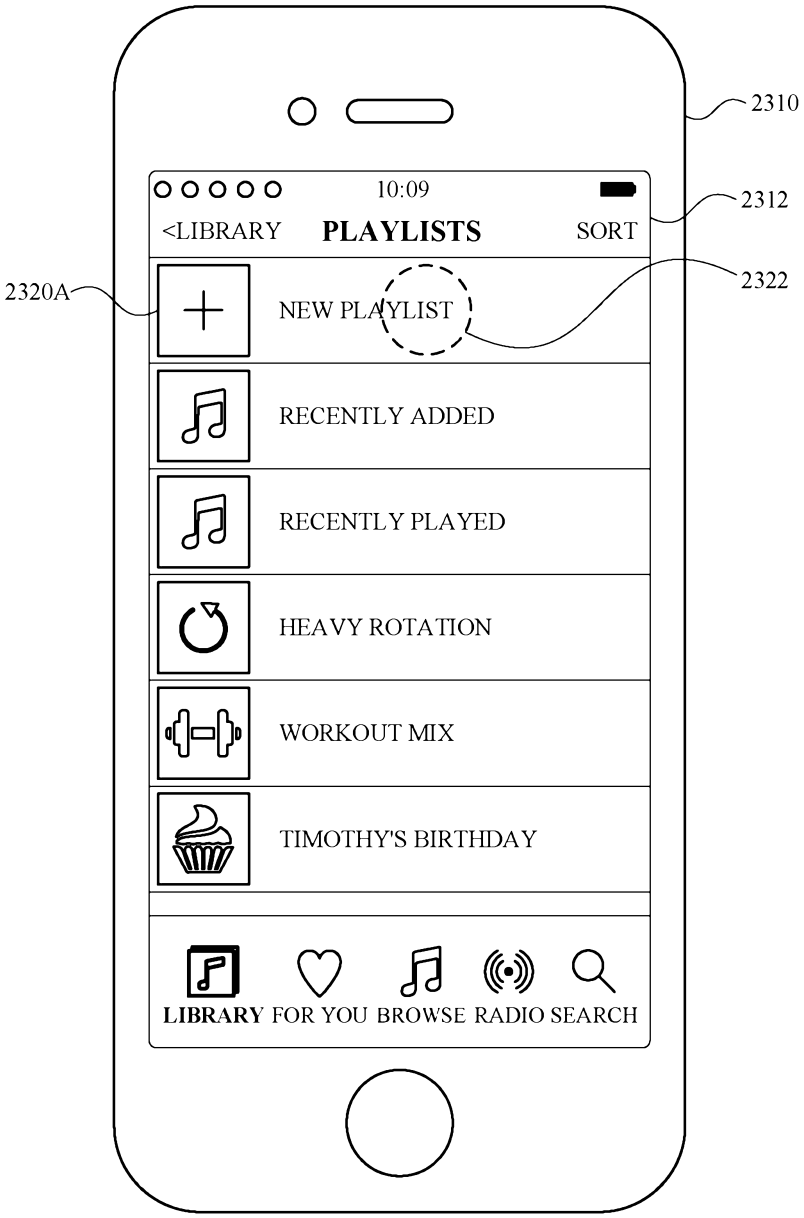

2310
10:09
2312
<LIBRARY
PLAYLISTS
SORT
2320A
2322
NEW PLAYLIST
RECENTLY ADDED
RECENTLY PLAYED
HEAVY ROTATION
WORKOUT MIX
TIMOTHY'S BIRTHDAY
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 23G

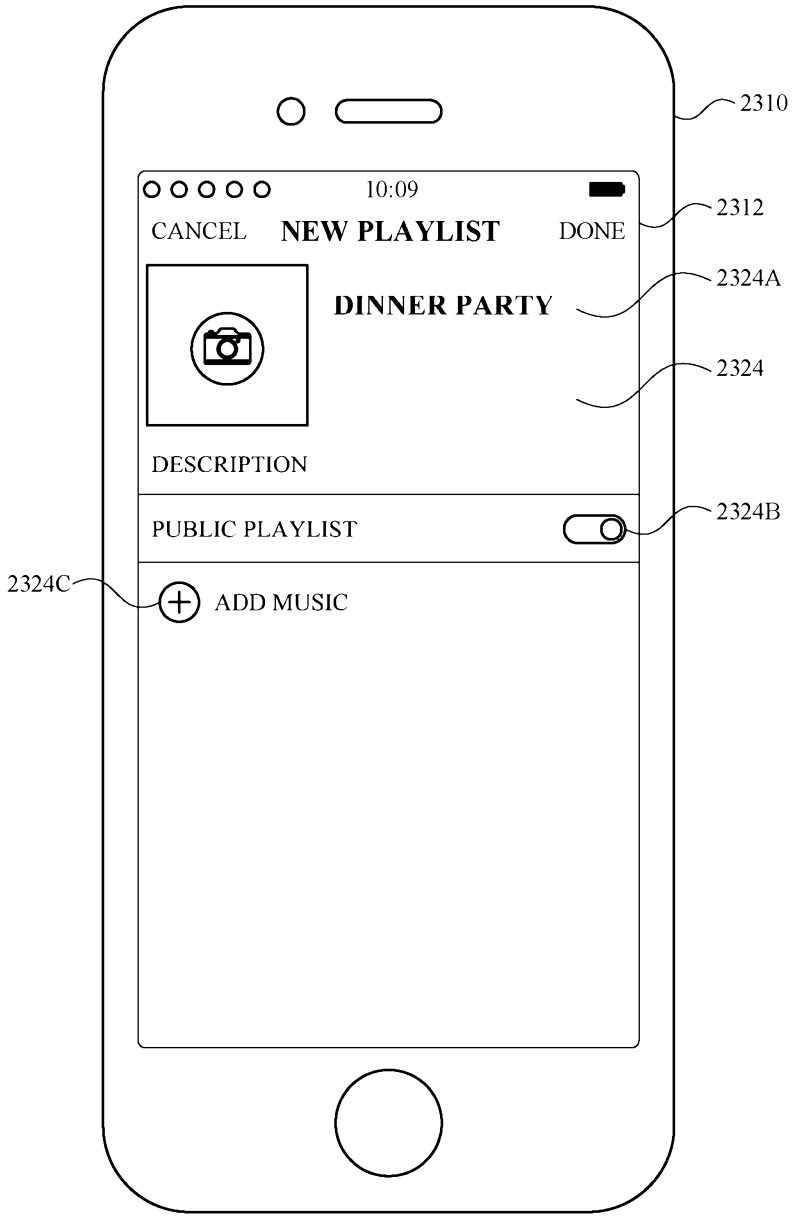

2310
10:09
CANCEL
NEW PLAYLIST
DONE
2312
2324A
DINNER PARTY
2324
DESCRIPTION
PUBLIC PLAYLIST
2324B
2324C
ADD MUSIC

FIG. 23H

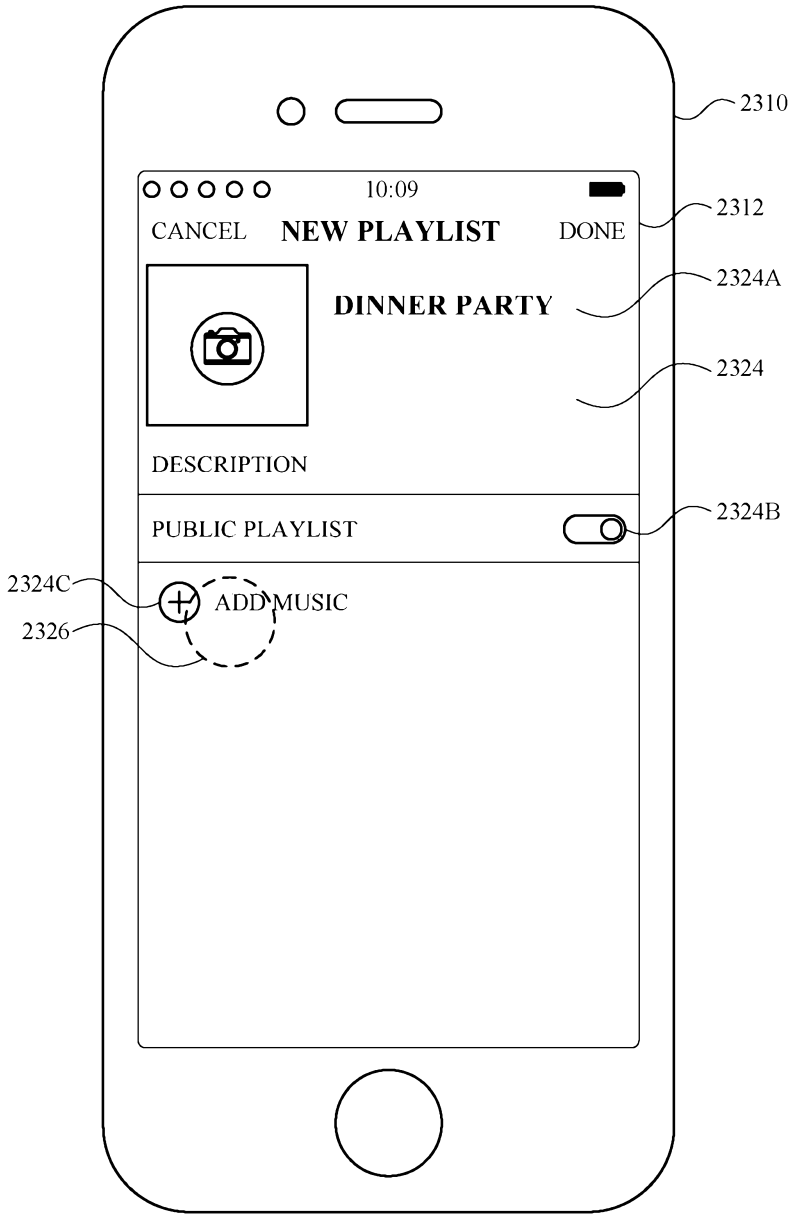

2310
10:09
CANCEL
NEW PLAYLIST
DONE
2312
2324A
DINNER PARTY
2324
DESCRIPTION
PUBLIC PLAYLIST
2324B
2324C
ADD MUSIC
2326

FIG. 23I

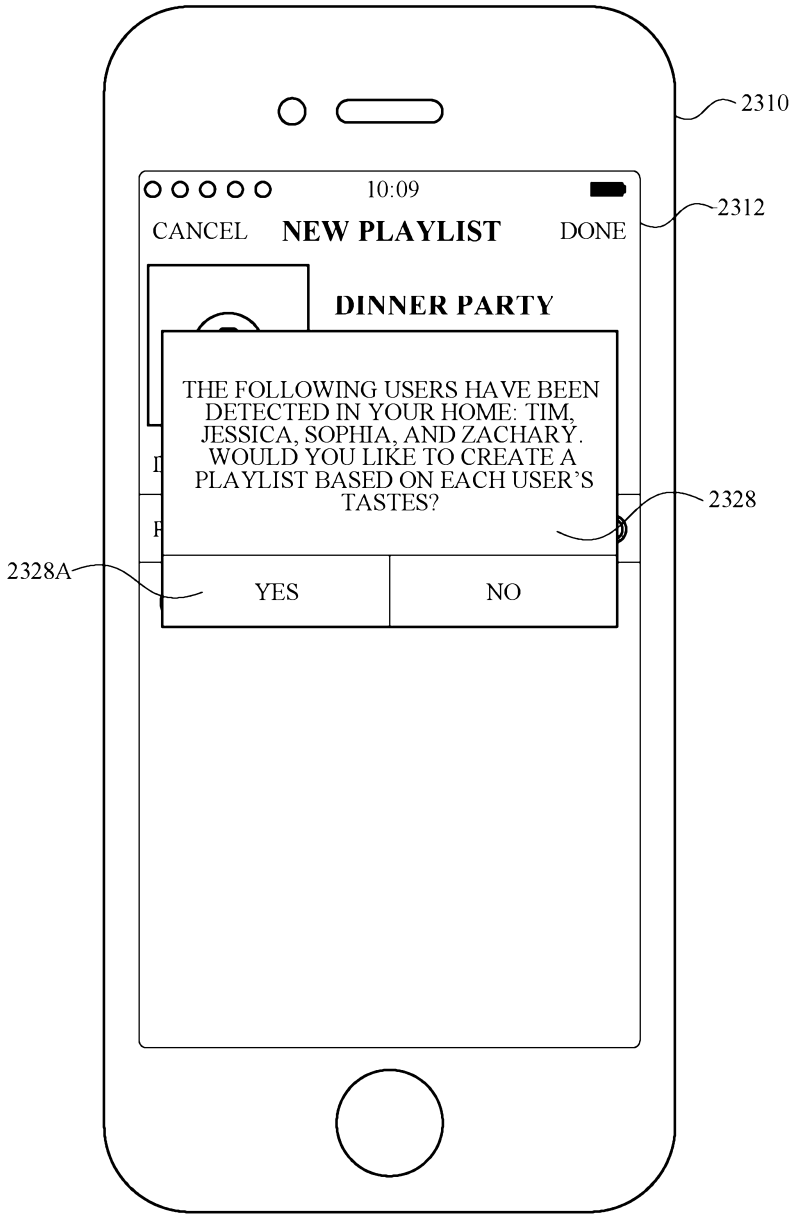

2310
10:09
2312
CANCEL
NEW PLAYLIST
DONE
DINNER PARTY
THE FOLLOWING USERS HAVE BEEN DETECTED IN YOUR HOME: TIM, JESSICA, SOPHIA, AND ZACHARY. WOULD YOU LIKE TO CREATE A PLAYLIST BASED ON EACH USER'S TASTES?
2328
2328A
YES
NO

FIG. 23J

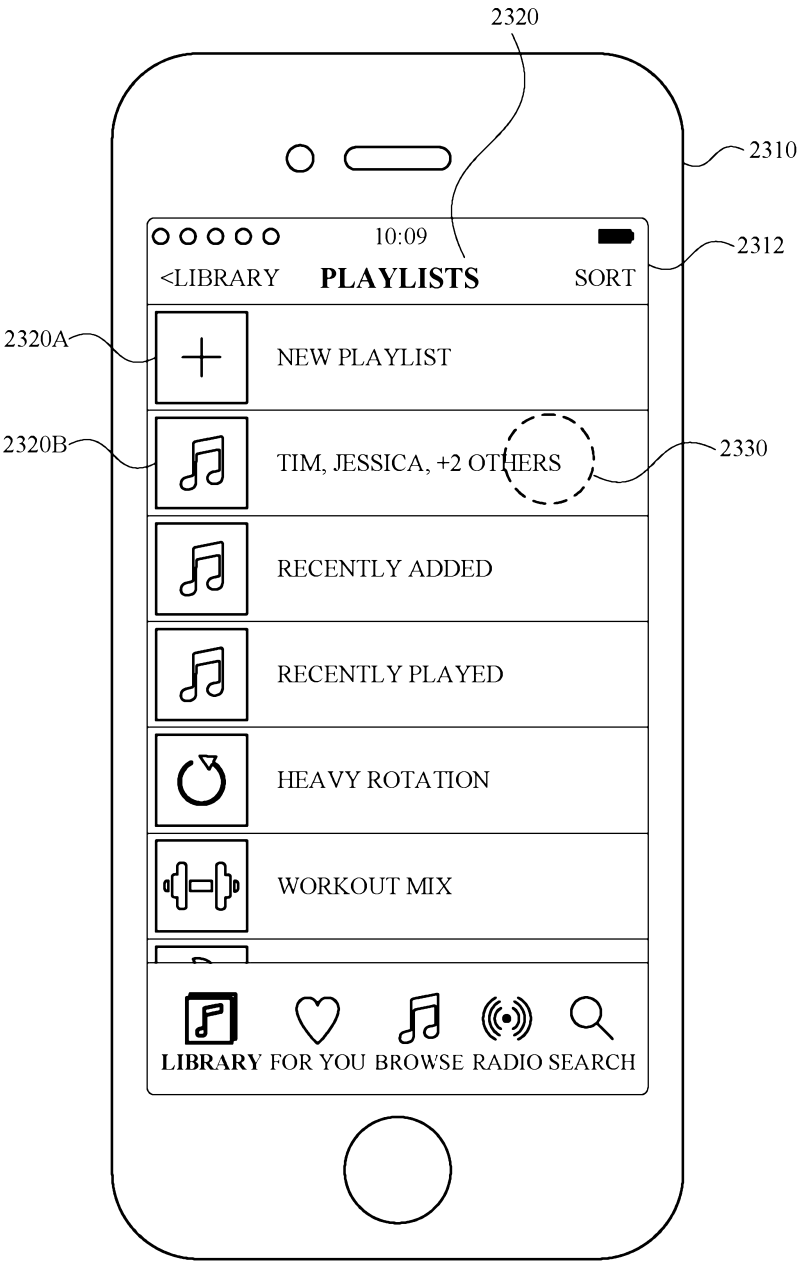

2320
2310
10:09
2312
<LIBRARY
PLAYLISTS
SORT
2320A
NEW PLAYLIST
2320B
TIM, JESSICA, +2 OTHERS
2330
RECENTLY ADDED
RECENTLY PLAYED
HEAVY ROTATION
WORKOUT MIX
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 23K

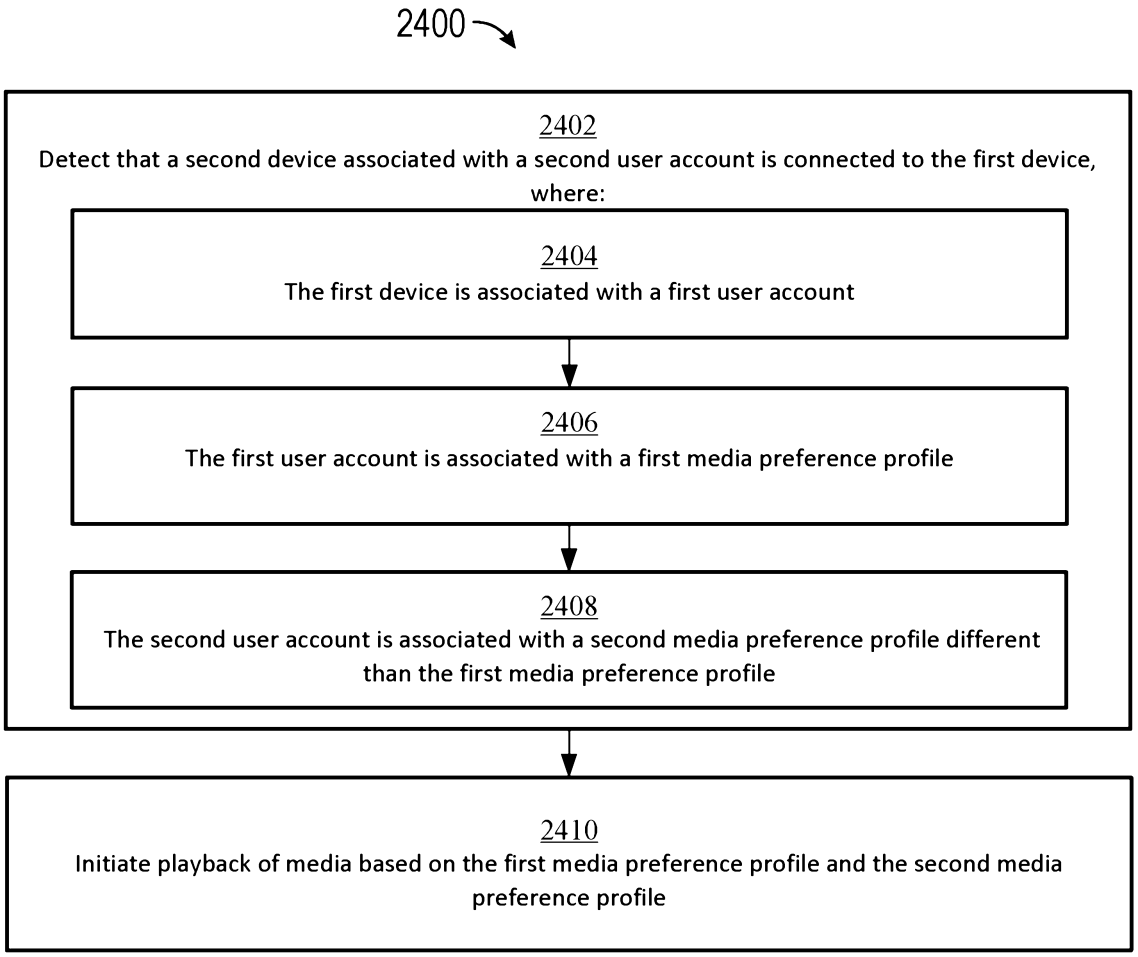

2400
2402
Detect that a second device associated with a second user account is connected to the first device, where:
2404
The first device is associated with a first user account
2406
The first user account is associated with a first media preference profile
2408
The second user account is associated with a second media preference profile different than the first media preference profile
2410
Initiate playback of media based on the first media preference profile and the second media preference profile

FIG. 24

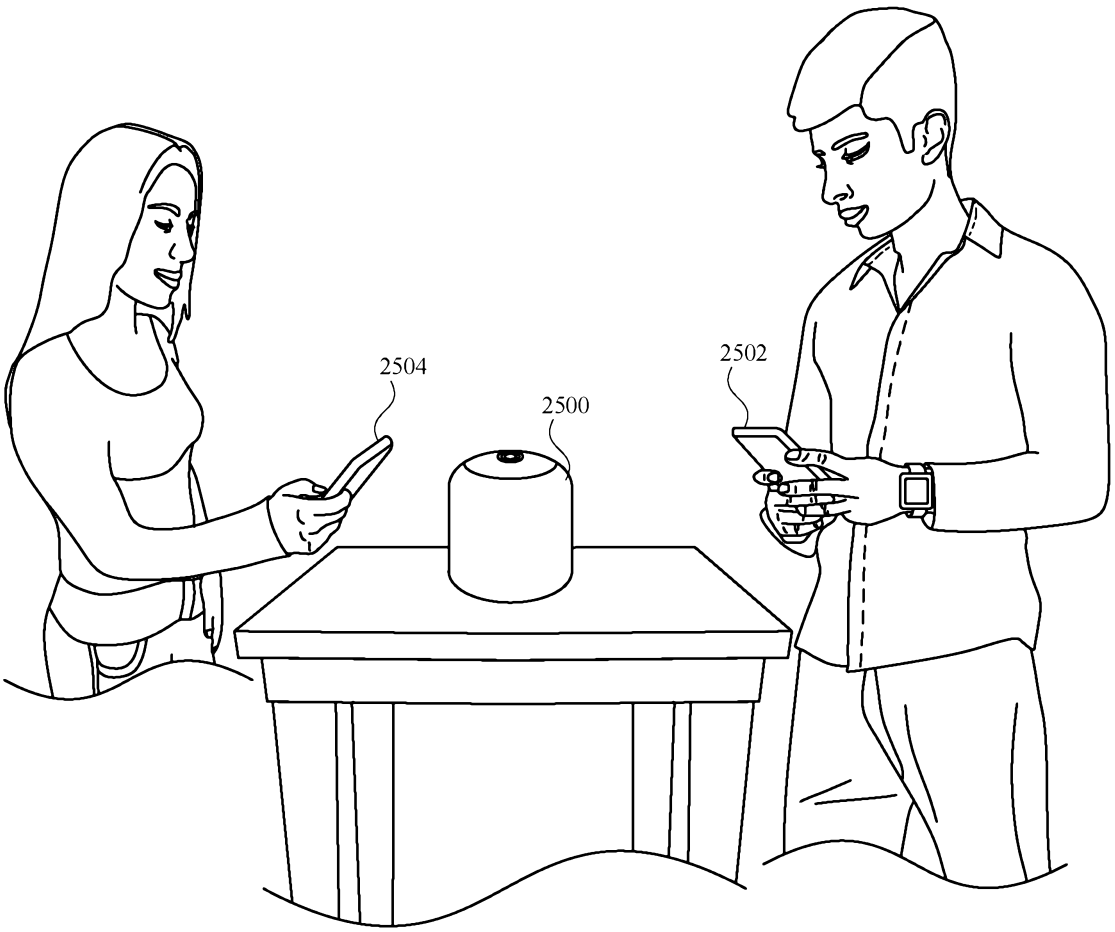


FIG. 25A

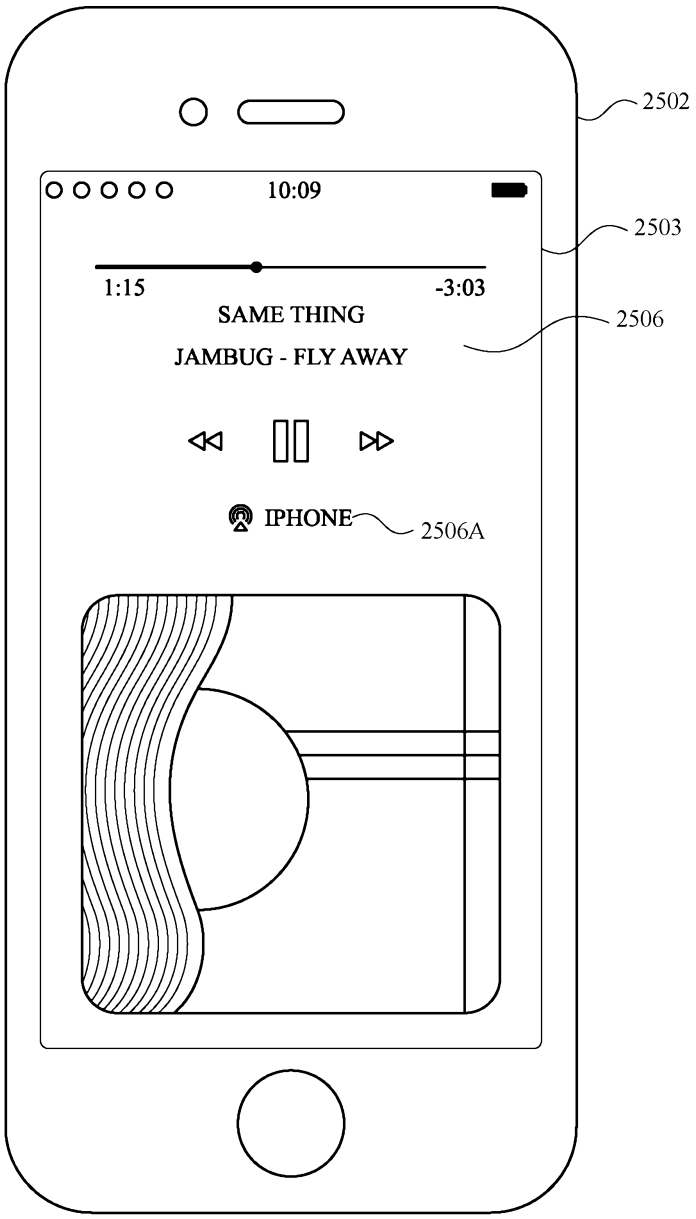

2502
10:09
2503
1:15
-3:03
SAME THING
2506
JAMBUG - FLY AWAY
IPHONE
2506A

FIG. 25B

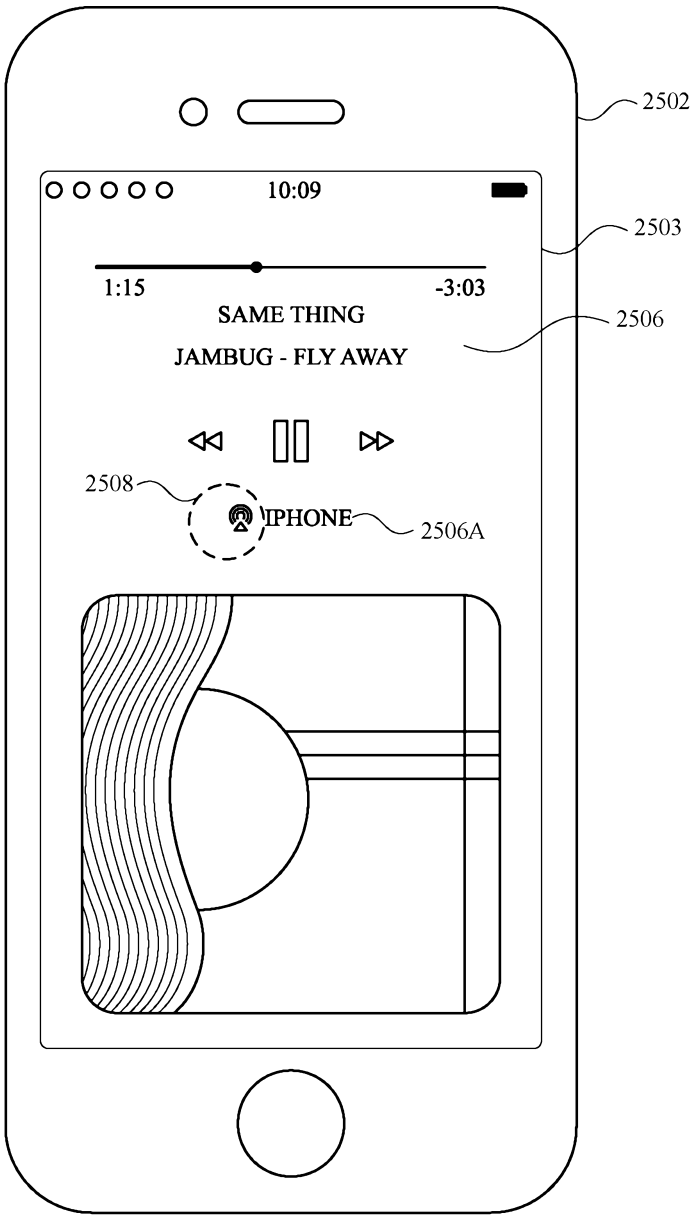

2502
10:09
2503
1:15
-3:03
SAME THING
2506
JAMBUG - FLY AWAY
2508
IPHONE
2506A

FIG. 25C

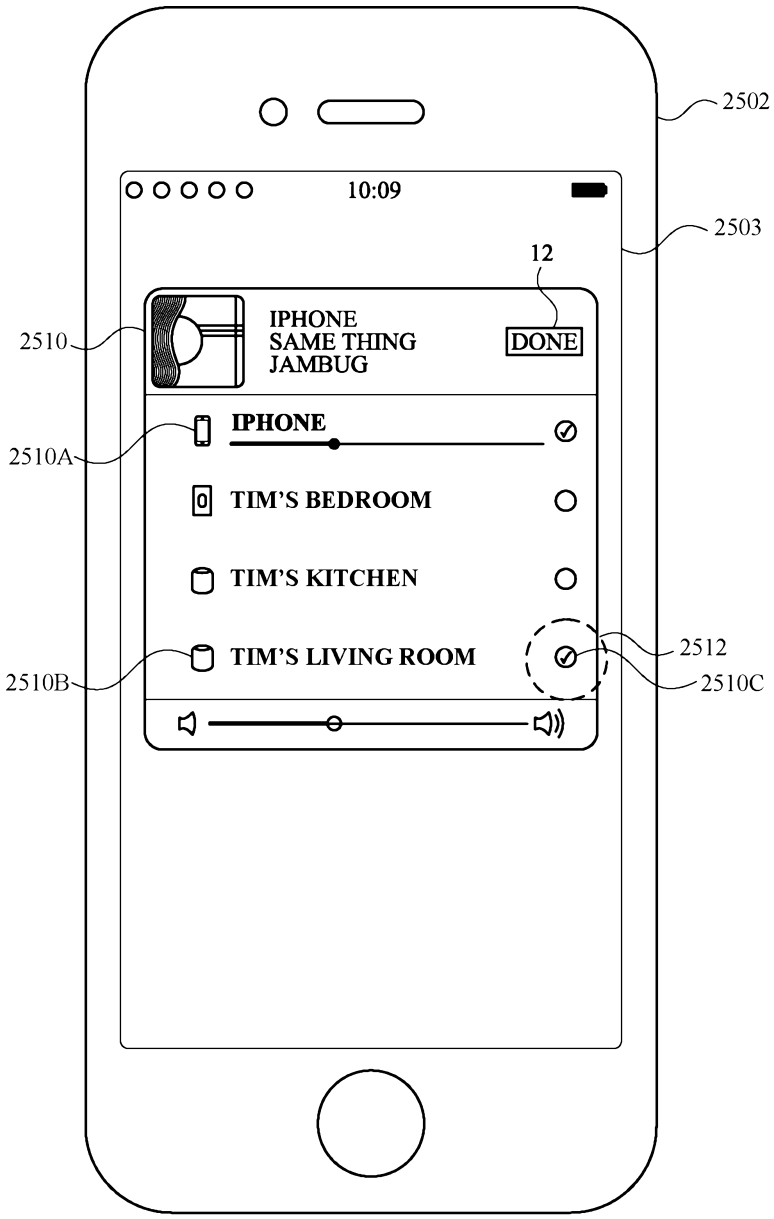

2502
10:09
2503
12
2510
IPHONE
SAME THING
JAMBUG
DONE
2510A
IPHONE
TIM'S BEDROOM
TIM'S KITCHEN
TIM'S LIVING ROOM
2512
2510B
2510C

FIG. 25D

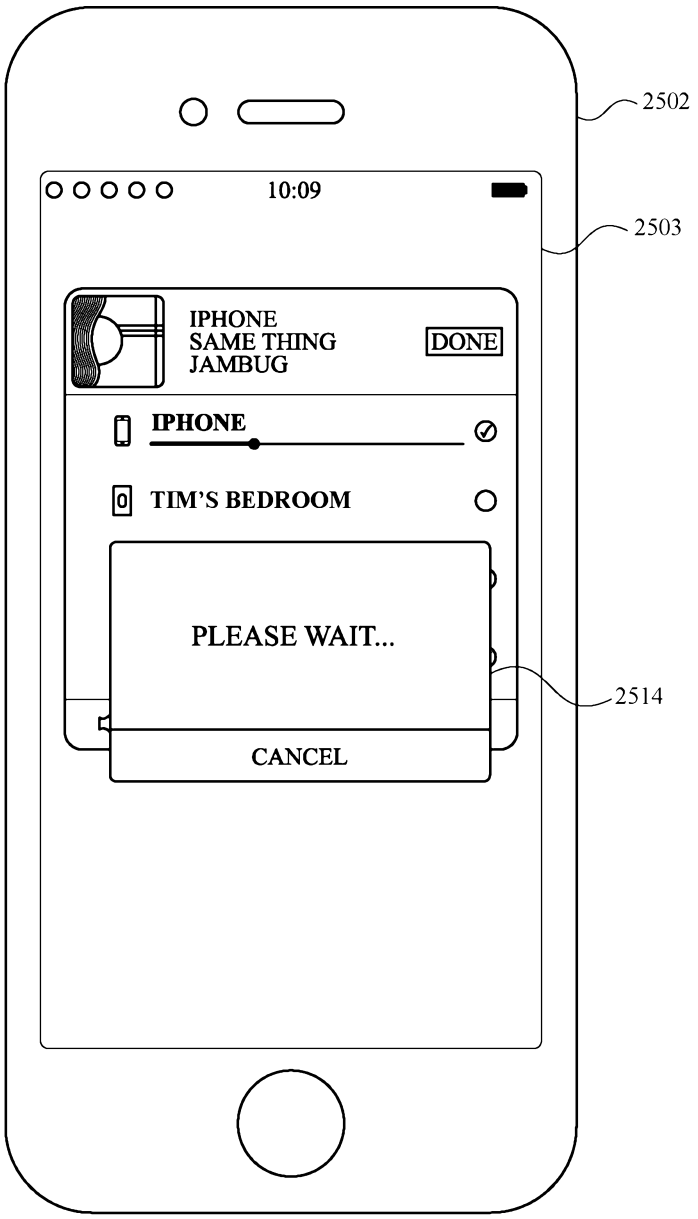

2502
10:09
2503
IPHONE
SAME THING
JAMBUG
DONE
IPHONE
TIM'S BEDROOM
PLEASE WAIT...
2514
CANCEL

FIG. 25E

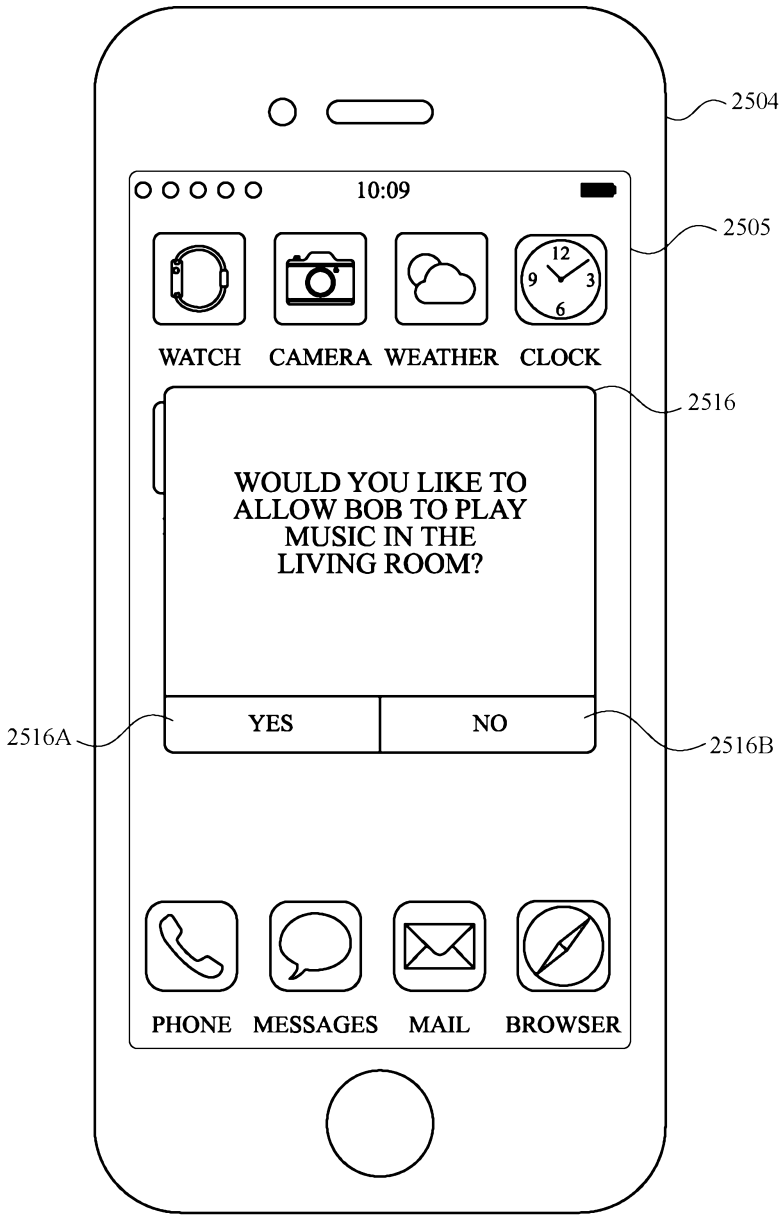

2504
10:09
2505
WATCH
CAMERA
WEATHER
CLOCK
2516
WOULD YOU LIKE TO ALLOW BOB TO PLAY MUSIC IN THE LIVING ROOM?
2516A
YES
NO
2516B
PHONE
MESSAGES
MAIL
BROWSER

FIG. 25F

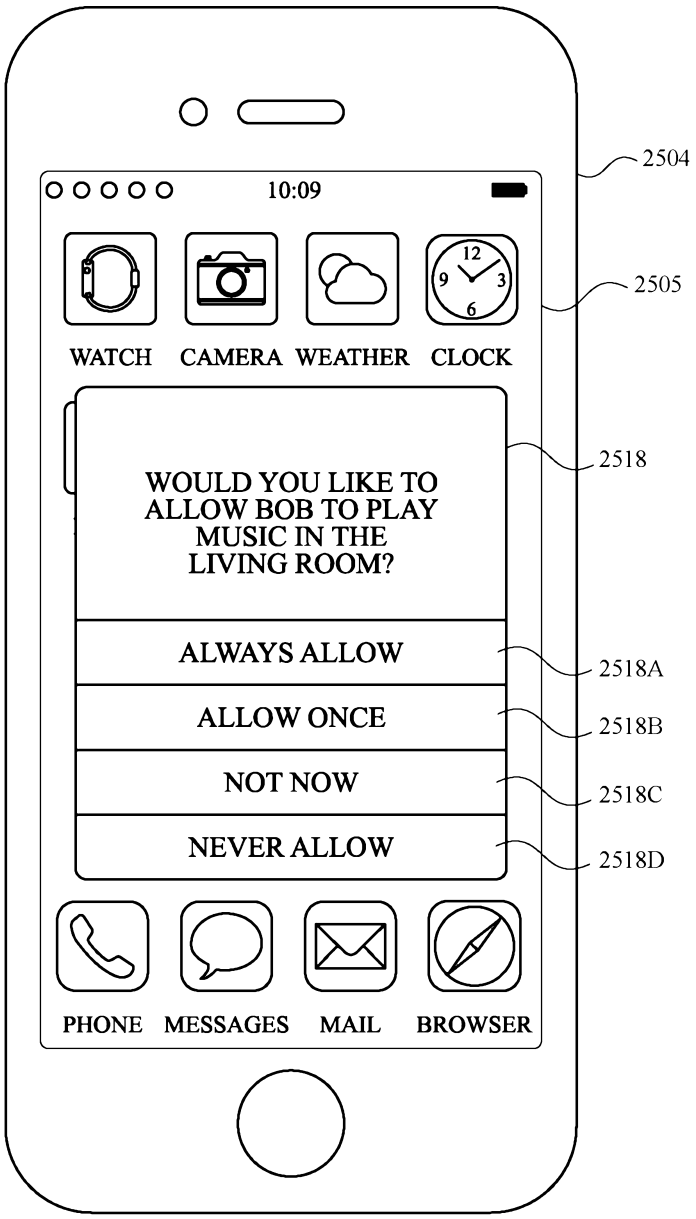

2504
10:09
2505
WATCH
CAMERA
WEATHER
CLOCK
2518
WOULD YOU LIKE TO ALLOW BOB TO PLAY MUSIC IN THE LIVING ROOM?
ALWAYS ALLOW
2518A
ALLOW ONCE
2518B
NOT NOW
2518C
NEVER ALLOW
2518D
PHONE
MESSAGES
MAIL
BROWSER

FIG. 25G

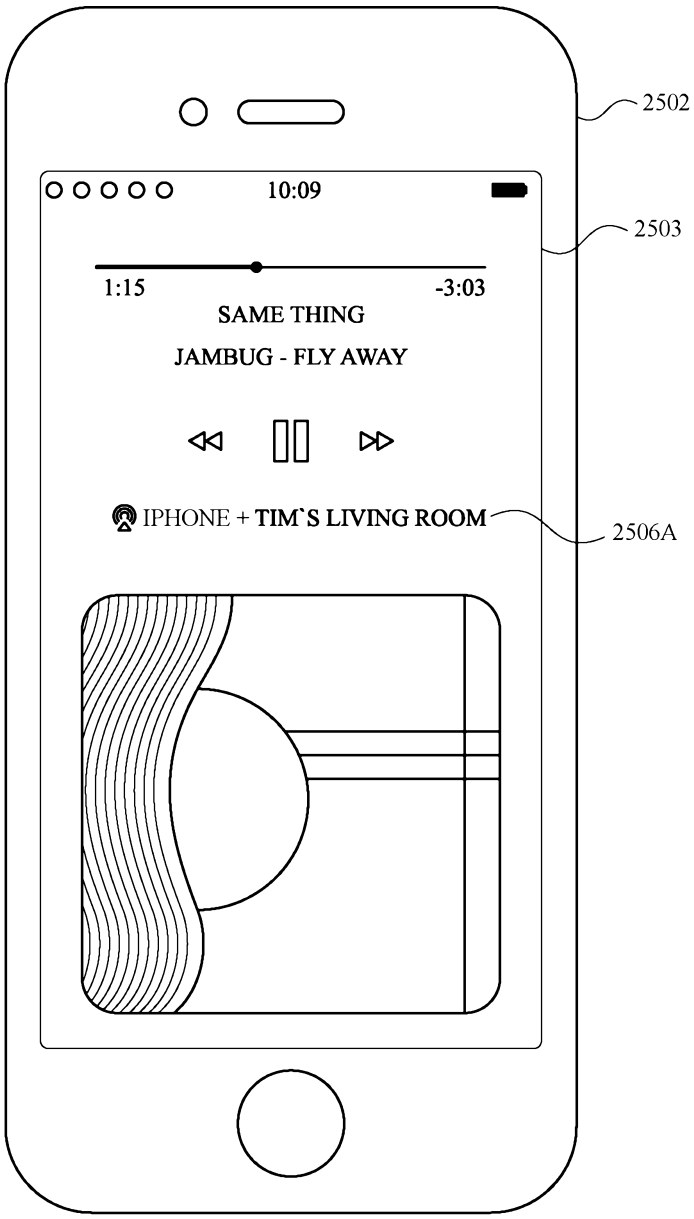

2502
10:09
2503
1:15
-3:03
SAME THING
JAMBUG - FLY AWAY
IPHONE + TIM`S LIVING ROOM
2506A

FIG. 25H

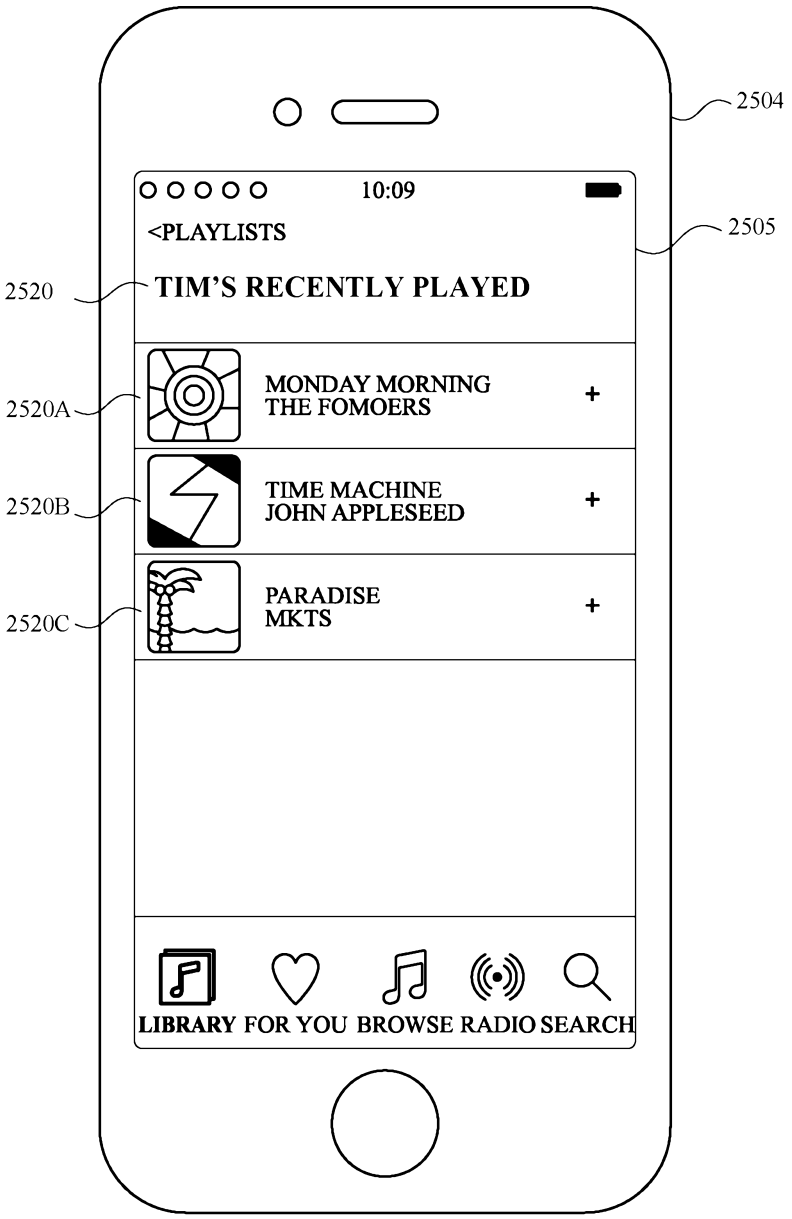

2504
10:09
<PLAYLISTS
2505
2520
TIM'S RECENTLY PLAYED
2520A
MONDAY MORNING
THE FOMOERS
+
2520B
TIME MACHINE
JOHN APPLESEED
+
2520C
PARADISE
MKTS
+
LIBRARY FOR YOU BROWSE RADIO SEARCH

FIG. 25I

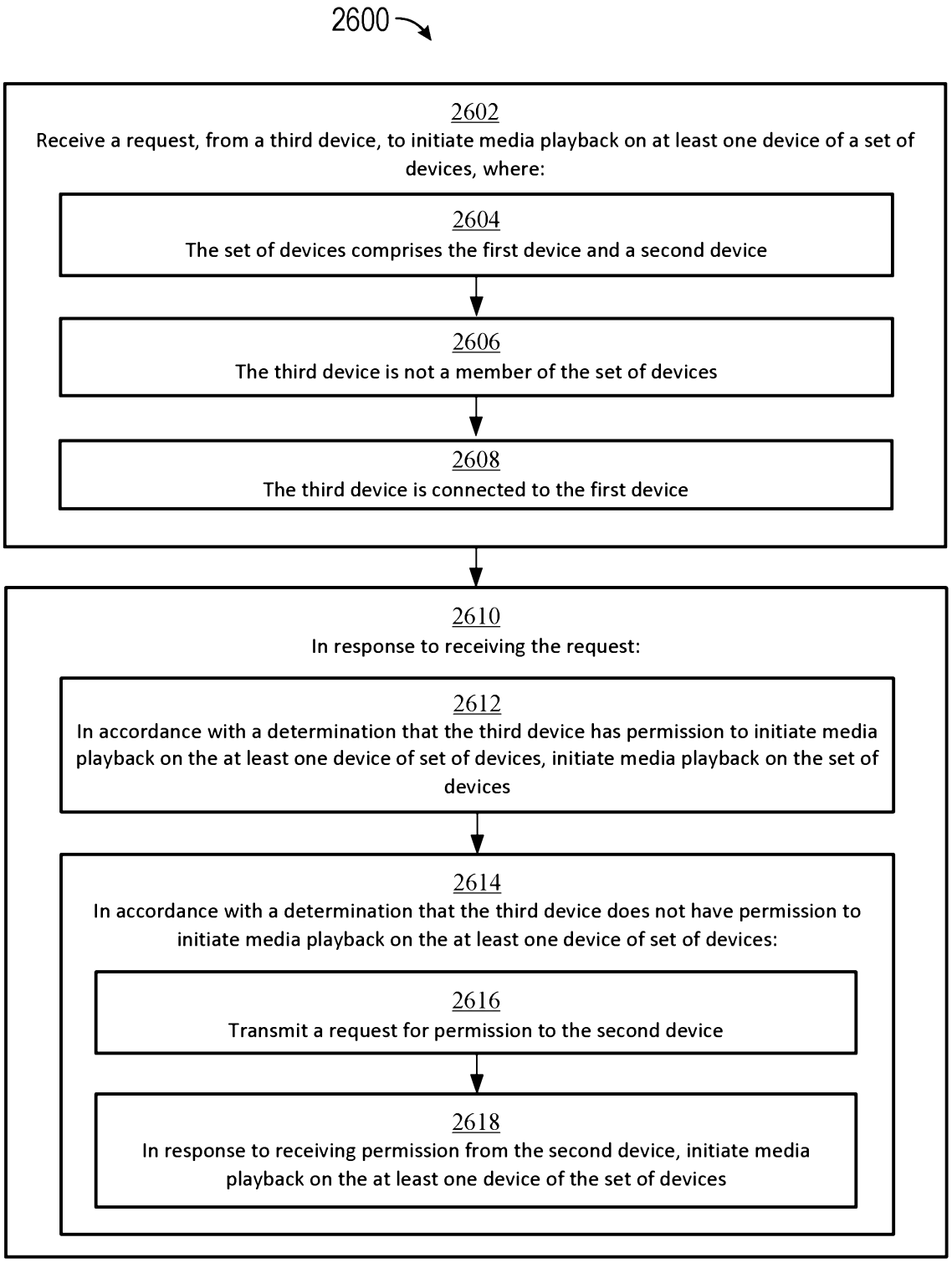

2600
2602
Receive a request, from a third device, to initiate media playback on at least one device of a set of devices, where:
2604
The set of devices comprises the first device and a second device
2606
The third device is not a member of the set of devices
2608
The third device is connected to the first device
2610
In response to receiving the request:
2612
In accordance with a determination that the third device has permission to initiate media playback on the at least one device of set of devices, initiate media playback on the set of devices
2614
In accordance with a determination that the third device does not have permission to initiate media playback on the at least one device of set of devices:
2616
Transmit a request for permission to the second device
2618
In response to receiving permission from the second device, initiate media playback on the at least one device of the set of devices

FIG. 26

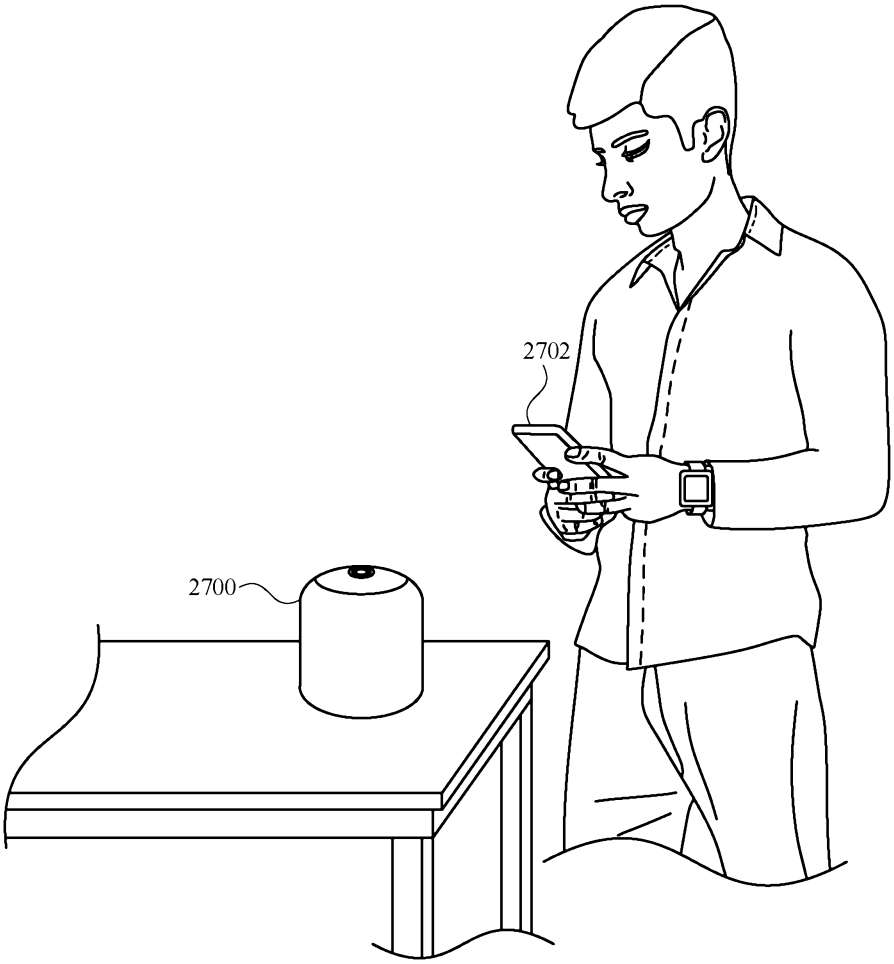


FIG. 27A

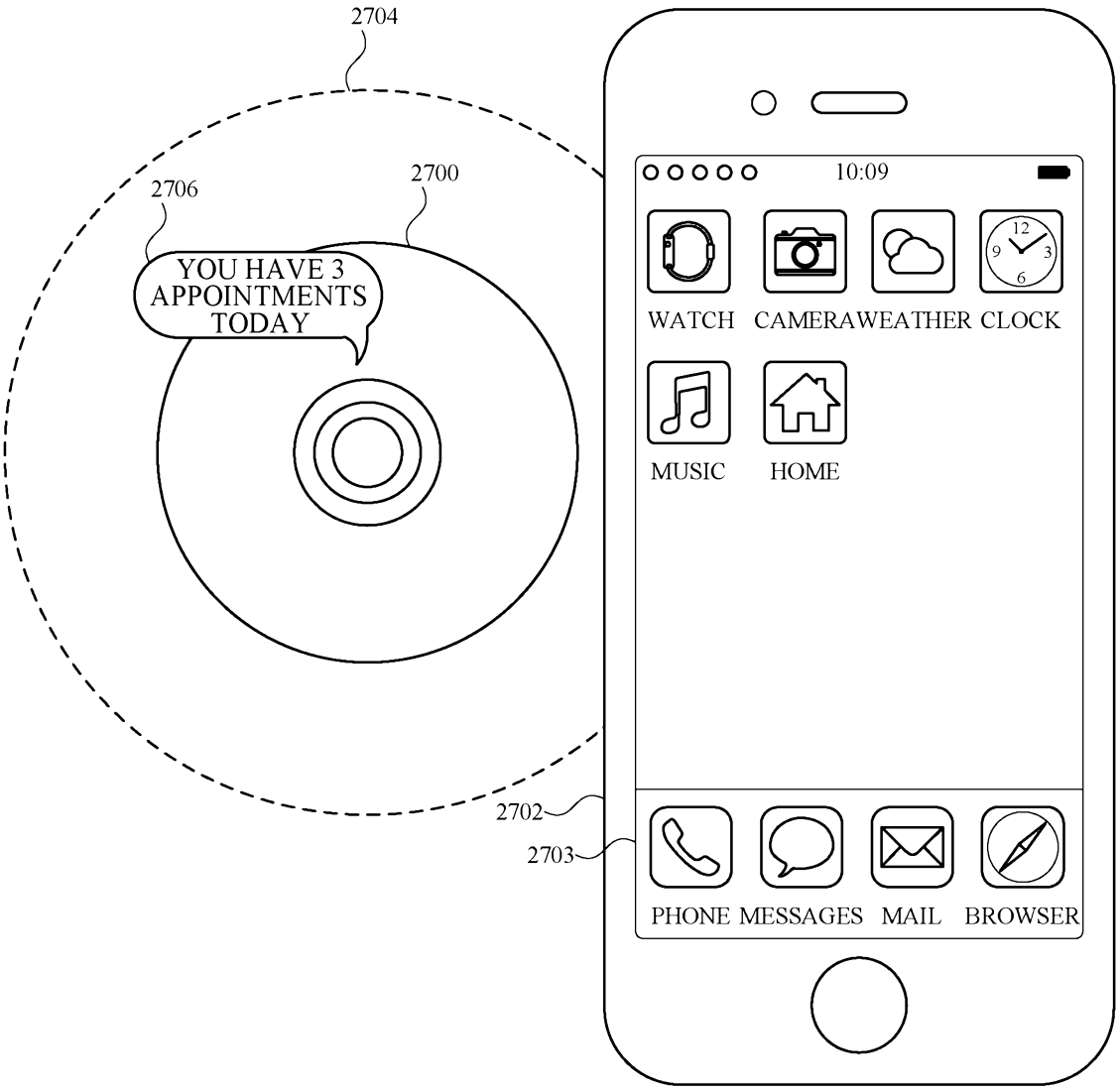

2704
2706
2700
YOU HAVE 3 APPOINTMENTS TODAY
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
2702
2703
PHONE
MESSAGES
MAIL
BROWSER

FIG. 27B

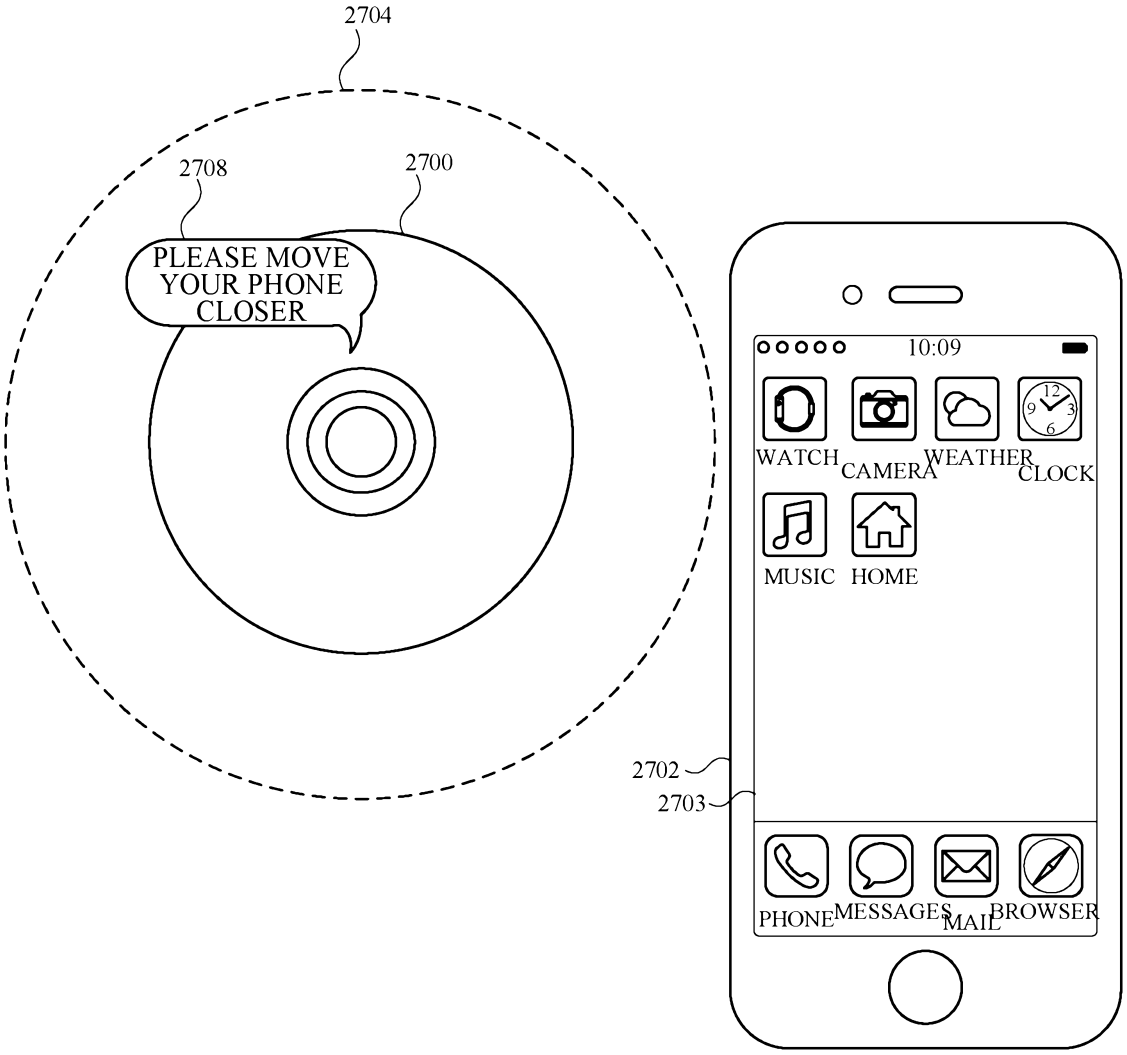

2704
2708
2700
PLEASE MOVE YOUR PHONE CLOSER
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
2702
2703
PHONE
MESSAGES
MAIL
BROWSER

FIG. 27C

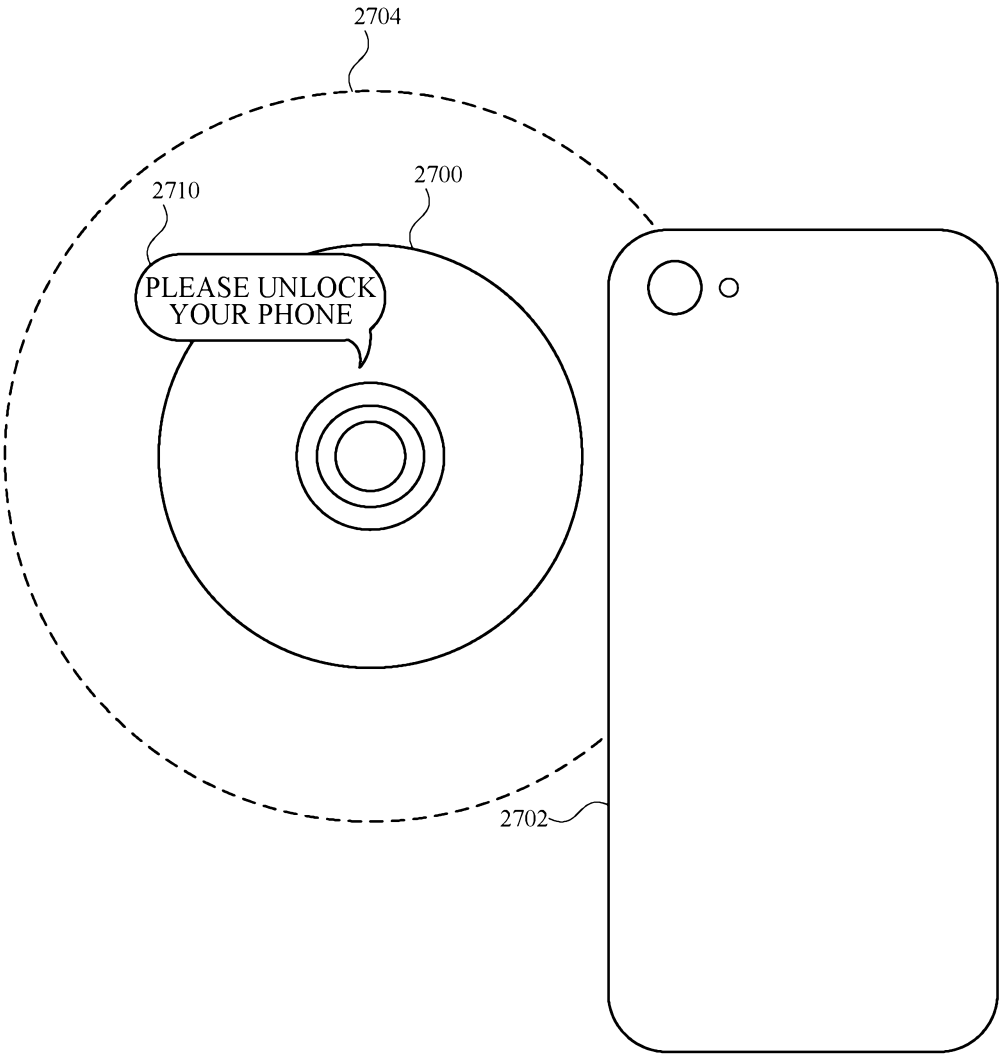


FIG. 27D

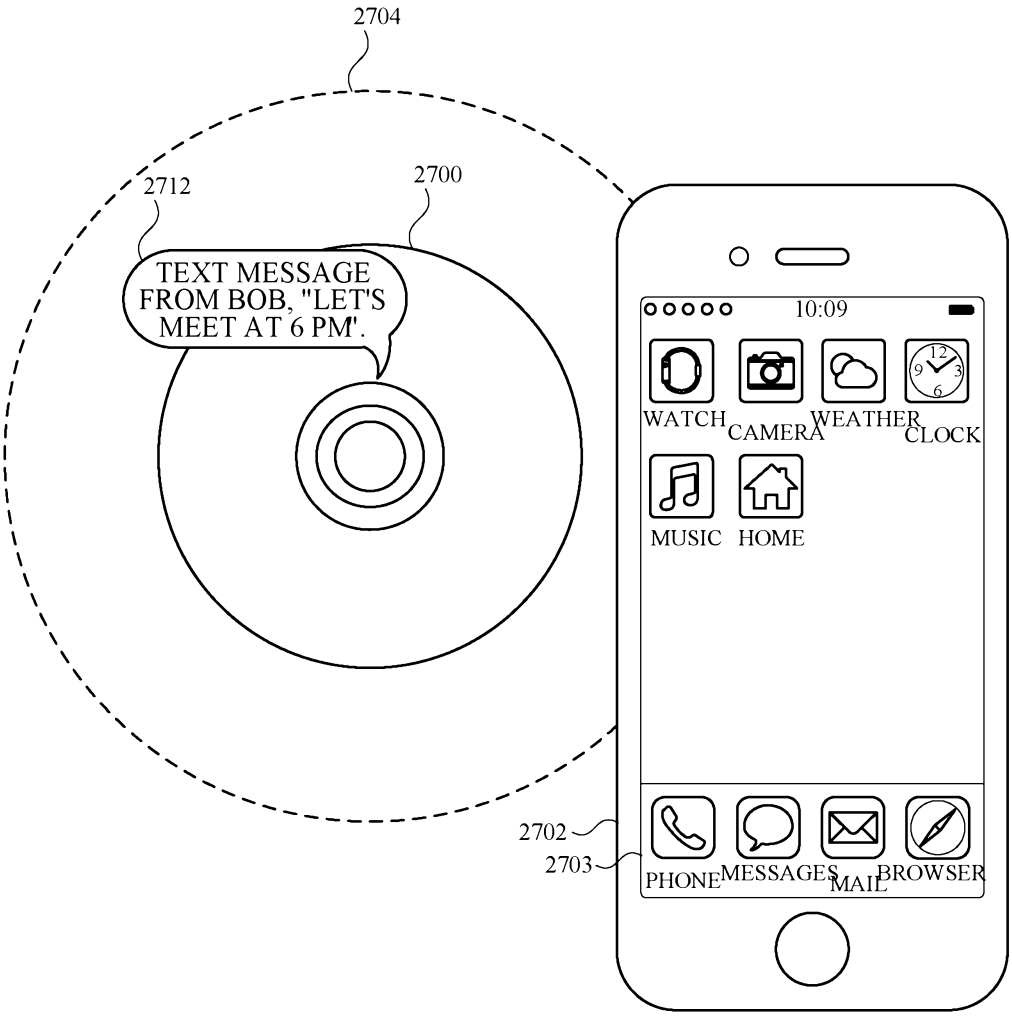

2704
2712
2700
TEXT MESSAGE FROM BOB, "LET'S MEET AT 6 PM'.
10:09
WATCH
CAMERA
WEATHER
CLOCK
MUSIC
HOME
2702
2703
PHONE
MESSAGES
MAIL
BROWSER

FIG. 27E

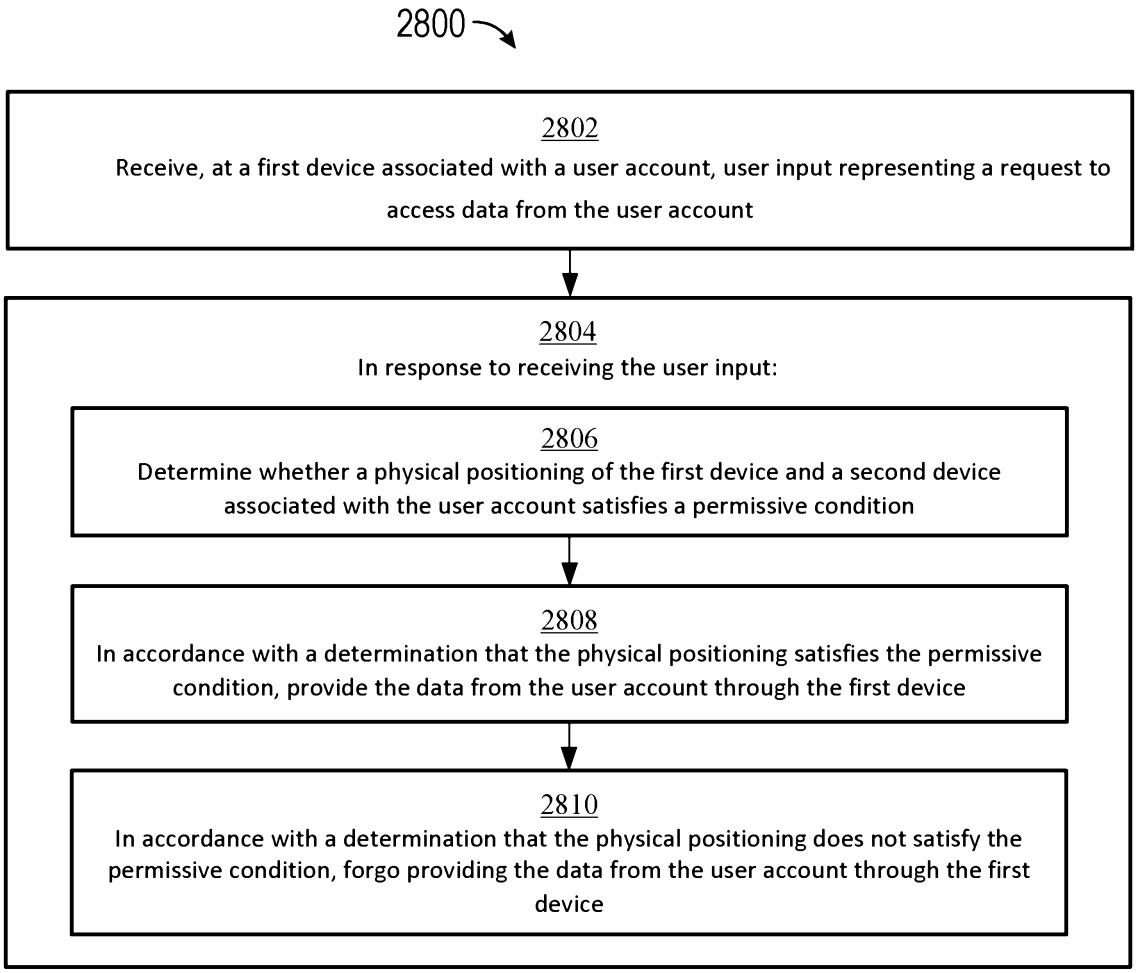

2800
2802
Receive, at a first device associated with a user account, user input representing a request to access data from the user account
2804
In response to receiving the user input:
2806
Determine whether a physical positioning of the first device and a second device associated with the user account satisfies a permissive condition
2808
In accordance with a determination that the physical positioning satisfies the permissive condition, provide the data from the user account through the first device
2810
In accordance with a determination that the physical positioning does not satisfy the permissive condition, forgo providing the data from the user account through the first device

FIG. 28

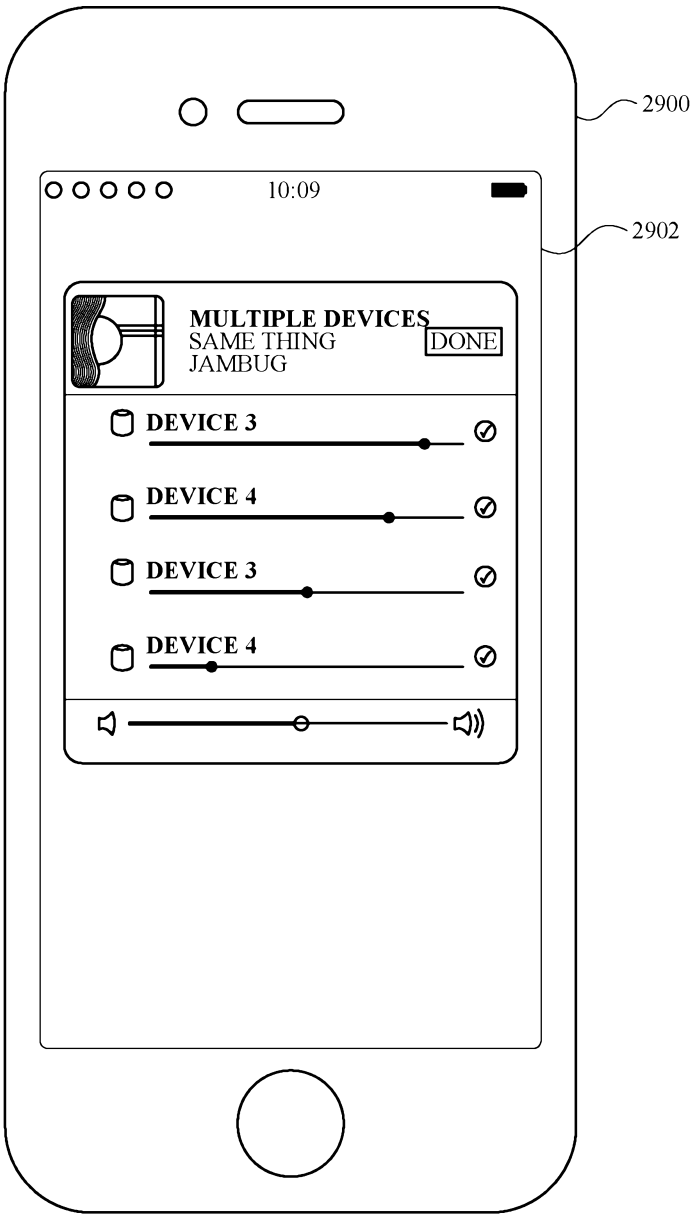

2900
10:09
2902
MULTIPLE DEVICES
SAME THING
JAMBUG
DONE
DEVICE 3
DEVICE 4
DEVICE 3
DEVICE 4

FIG. 29A

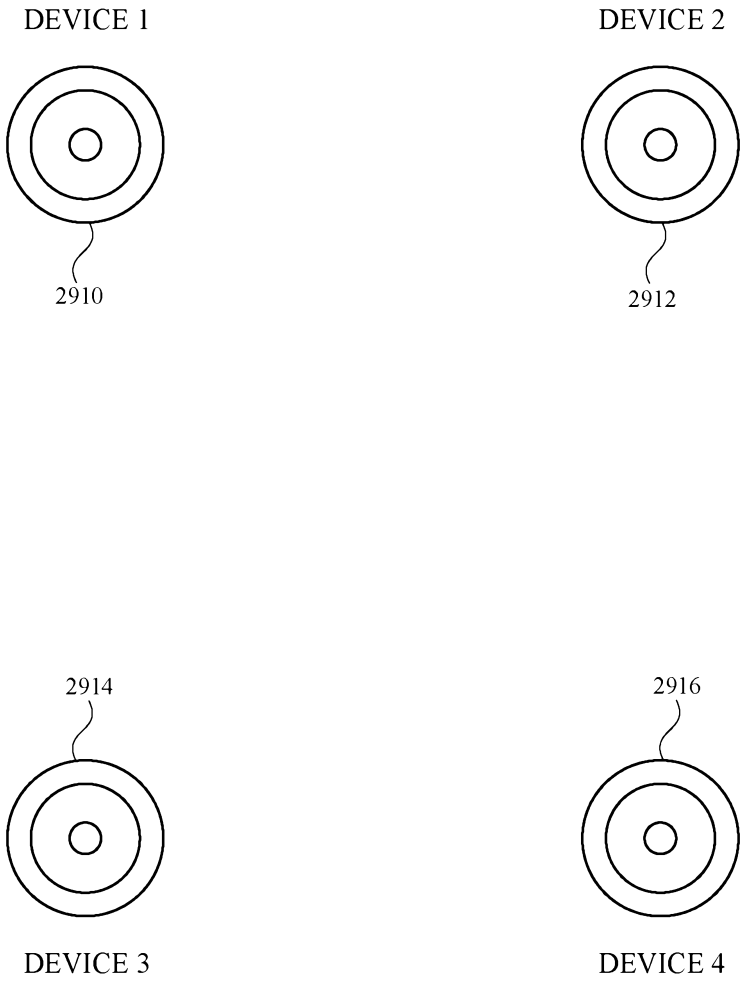


FIG. 29B

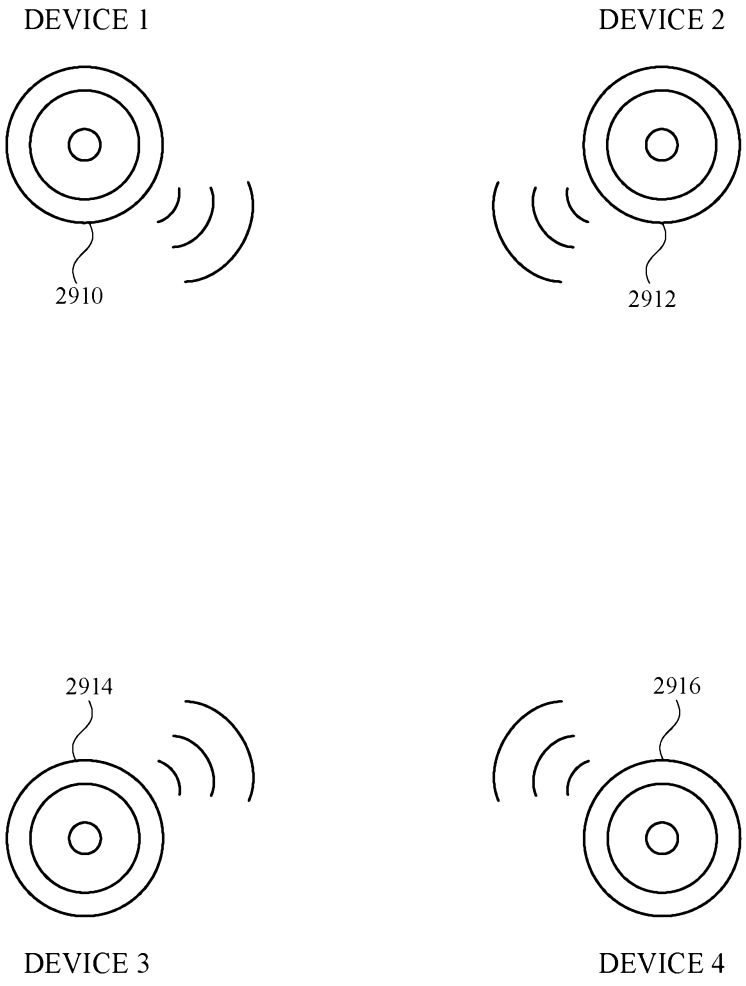


FIG. 29C

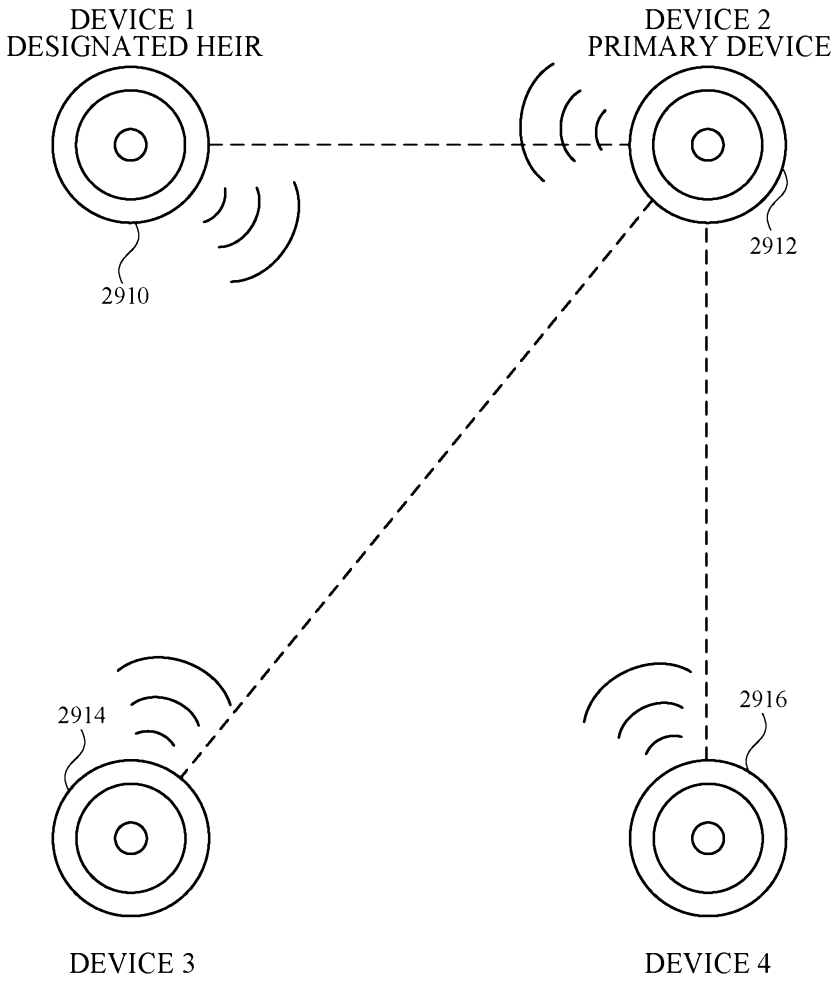

DEVICE 1
DESIGNATED HEIR
DEVICE 2
PRIMARY DEVICE
2910
2912
2914
2916
DEVICE 3
DEVICE 4

FIG. 29D

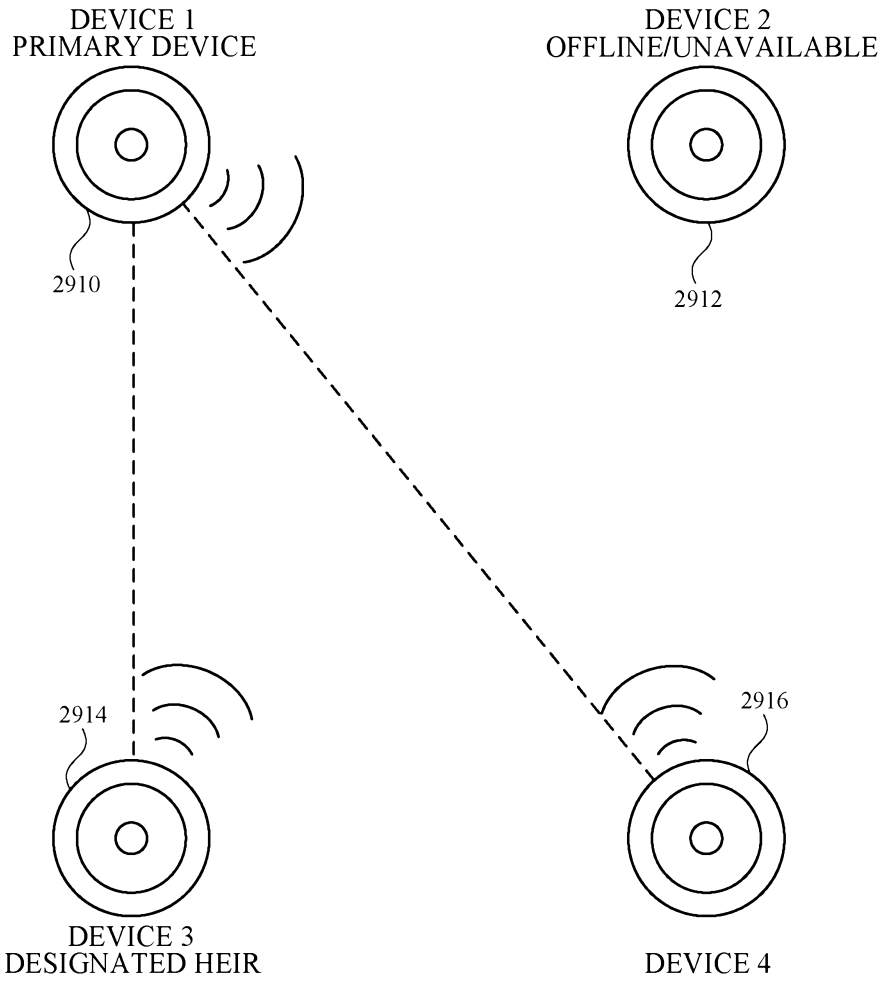

DEVICE 1
PRIMARY DEVICE
DEVICE 2
OFFLINE/UNAVAILABLE
2910
2912
2914
2916
DEVICE 3
DESIGNATED HEIR
DEVICE 4

FIG. 29E

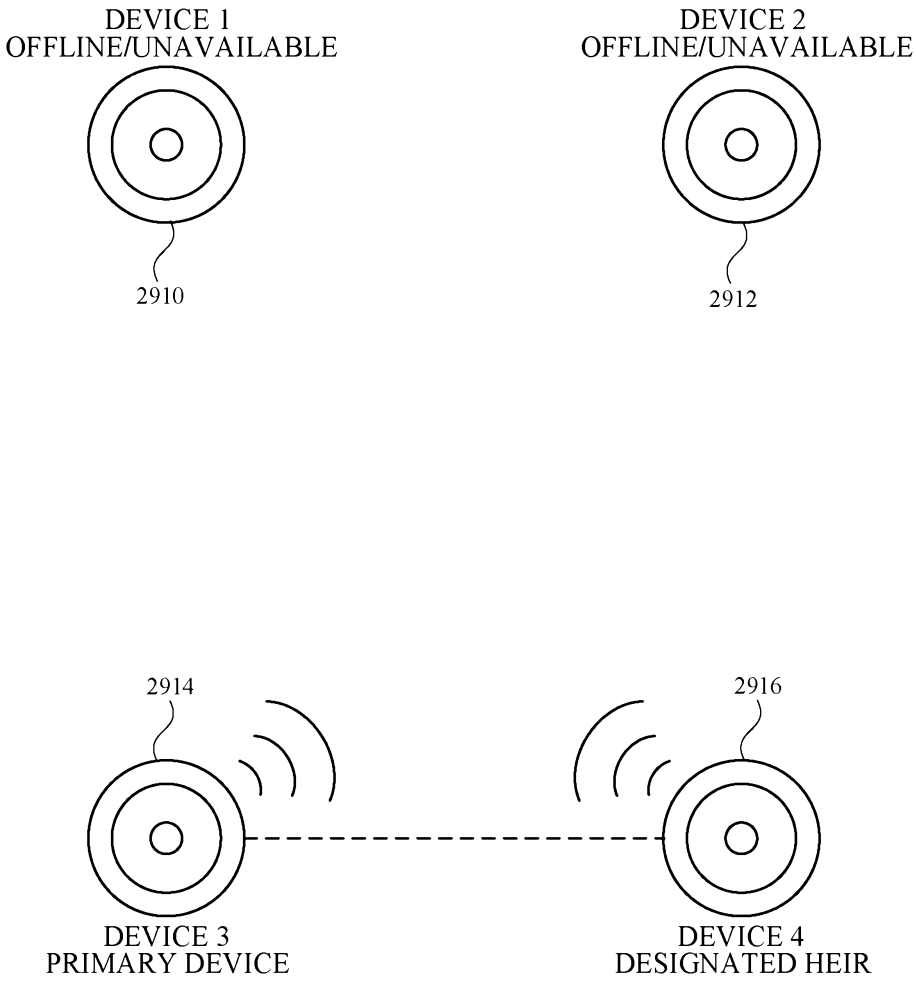

DEVICE 1
OFFLINE/UNAVAILABLE
2910
DEVICE 2
OFFLINE/UNAVAILABLE
2912
2914
2916
DEVICE 3
PRIMARY DEVICE
DEVICE 4
DESIGNATED HEIR

FIG. 29F

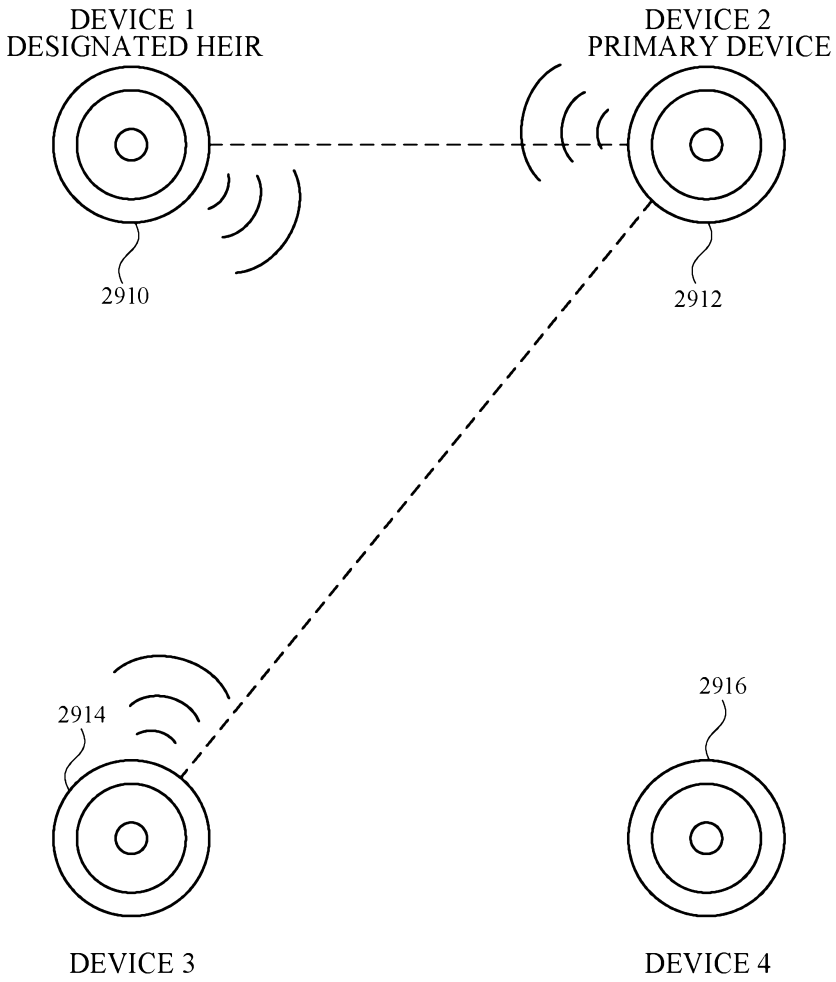

DEVICE 1
DESIGNATED HEIR
DEVICE 2
PRIMARY DEVICE
2910
2912
2914
2916
DEVICE 3
DEVICE 4

FIG. 29G

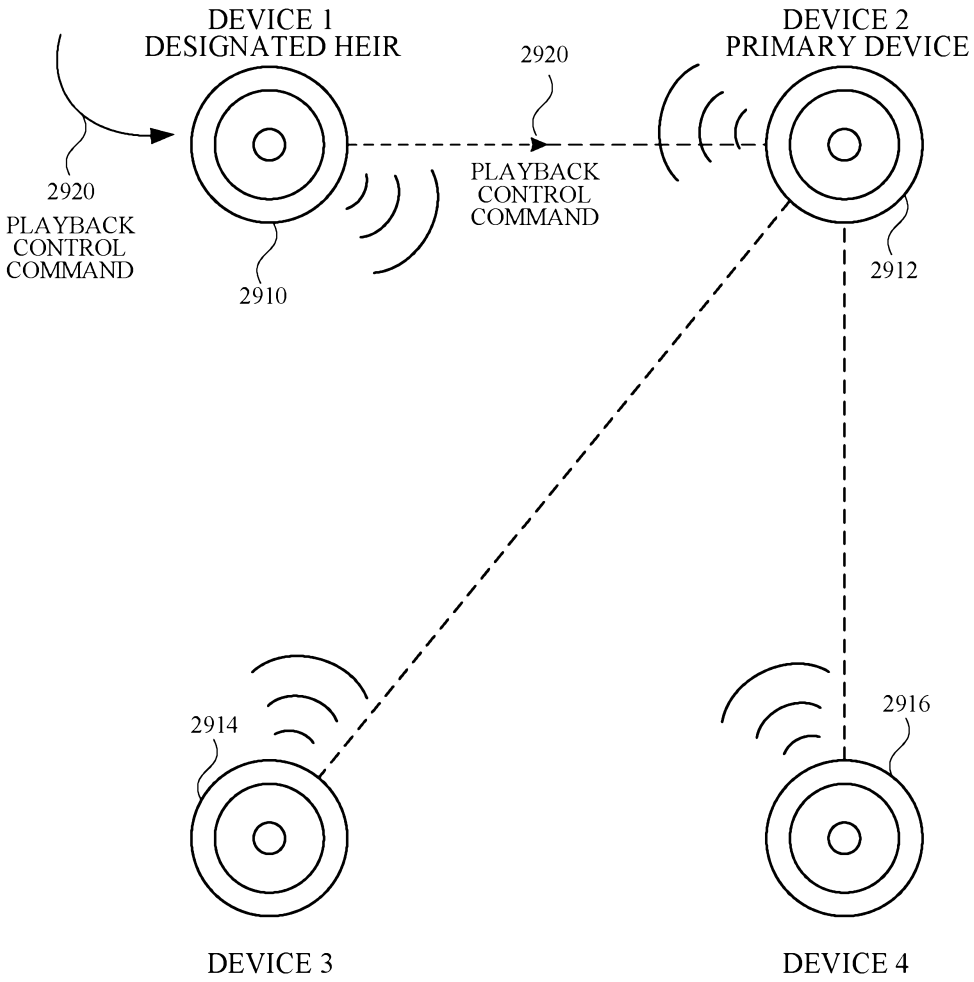

DEVICE 1
DESIGNATED HEIR
2920
PLAYBACK
CONTROL
COMMAND
2910
2920
PLAYBACK
CONTROL
COMMAND
DEVICE 2
PRIMARY DEVICE
2912
2914
2916
DEVICE 3
DEVICE 4

FIG. 29H

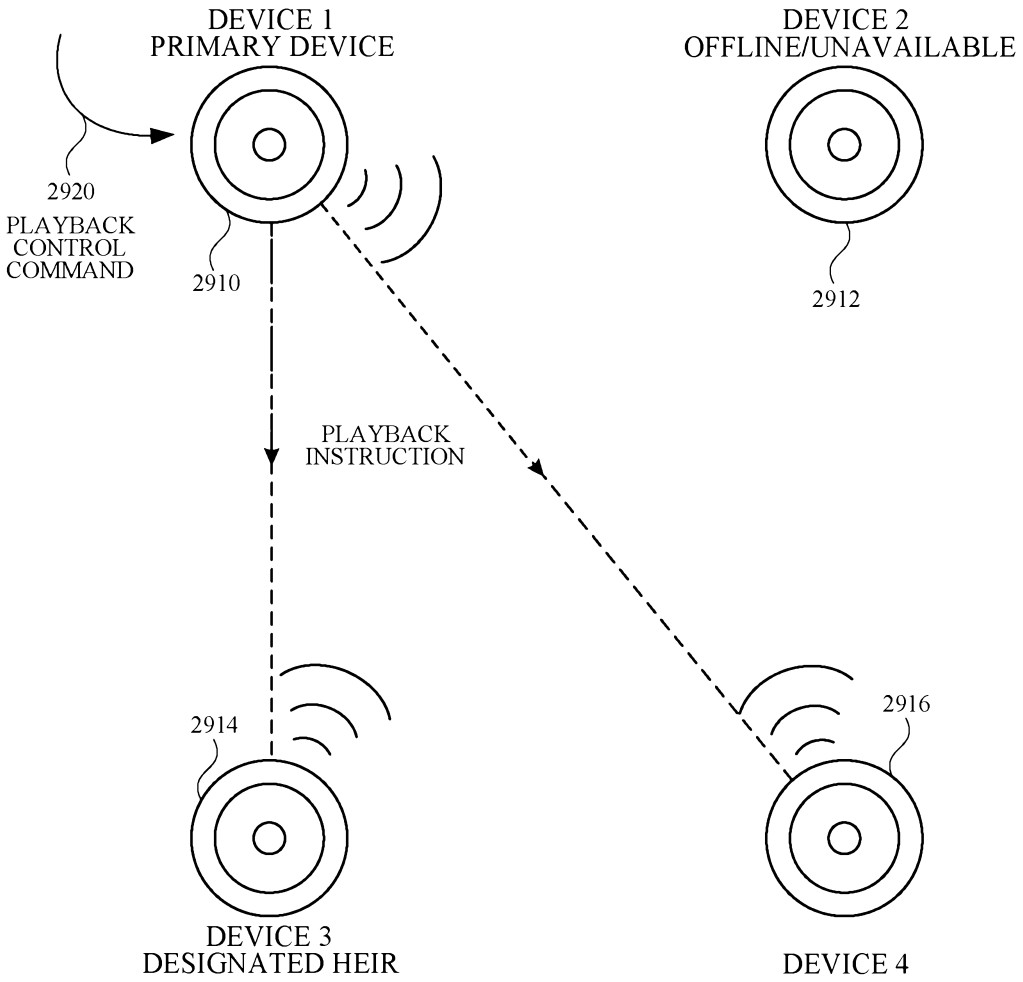

DEVICE 1
PRIMARY DEVICE
DEVICE 2
OFFLINE/UNAVAILABLE
2920
PLAYBACK
CONTROL
COMMAND
2910
2912
PLAYBACK
INSTRUCTION
2914
2916
DEVICE 3
DESIGNATED HEIR
DEVICE 4

FIG. 29I

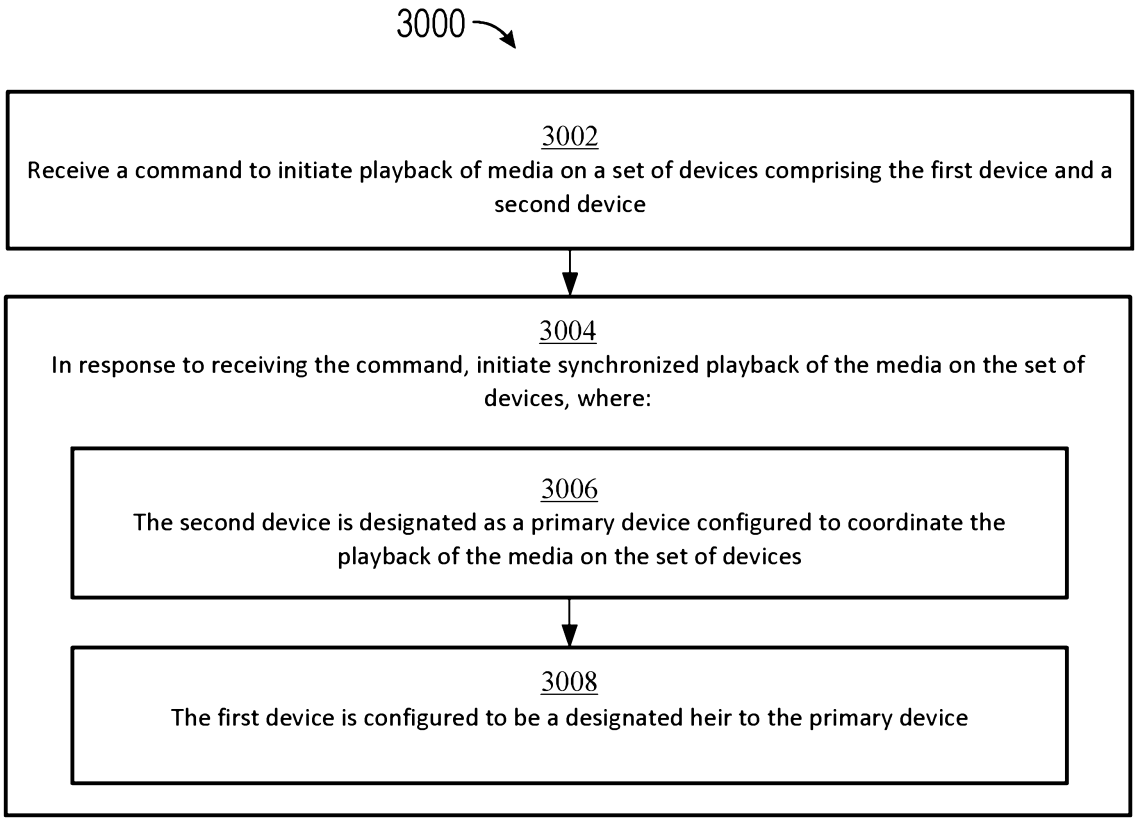

3000
3002
Receive a command to initiate playback of media on a set of devices comprising the first device and a second device
3004
In response to receiving the command, initiate synchronized playback of the media on the set of devices, where:
3006
The second device is designated as a primary device configured to coordinate the playback of the media on the set of devices
3008
The first device is configured to be a designated heir to the primary device

FIG. 30

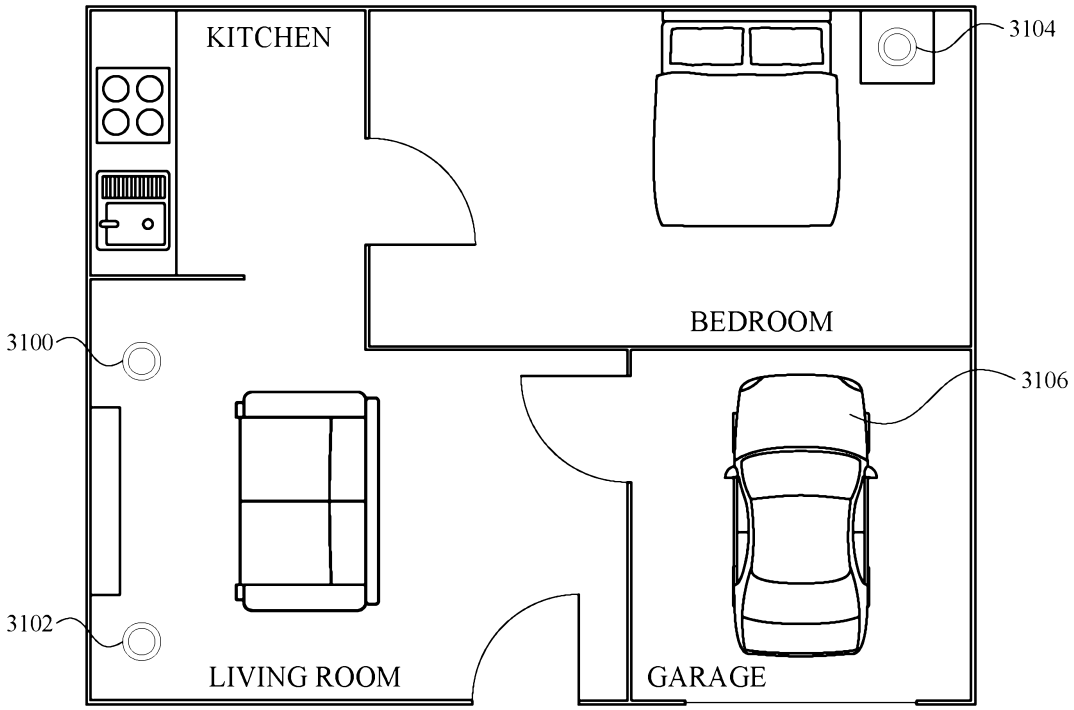


FIG. 31A

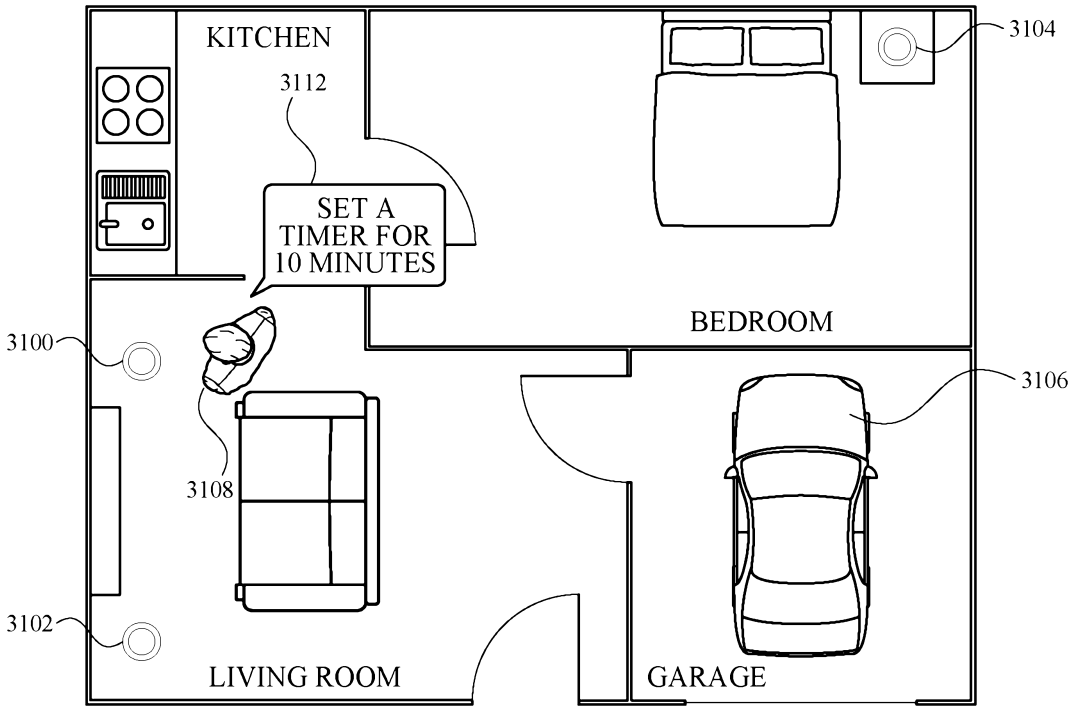

KITCHEN
3112
SET A TIMER FOR 10 MINUTES
3104
BEDROOM
3100
3106
3108
3102
LIVING ROOM
GARAGE

FIG. 31B

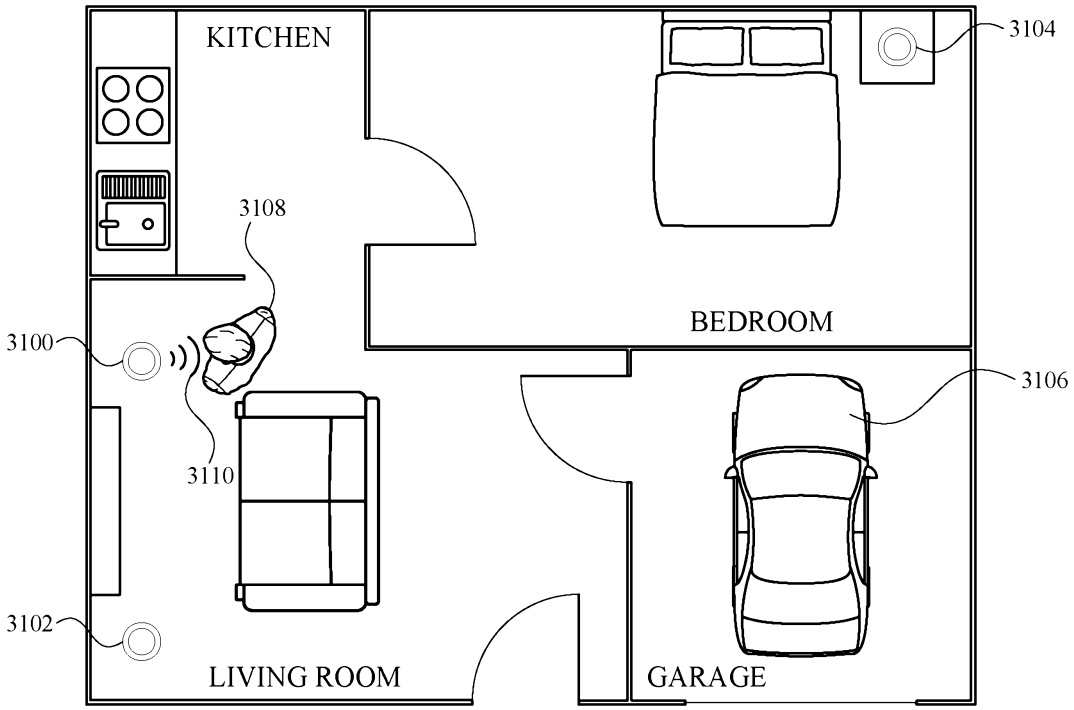

KITCHEN
3104
3108
BEDROOM
3100
3106
3110
3102
LIVING ROOM
GARAGE

FIG. 31C

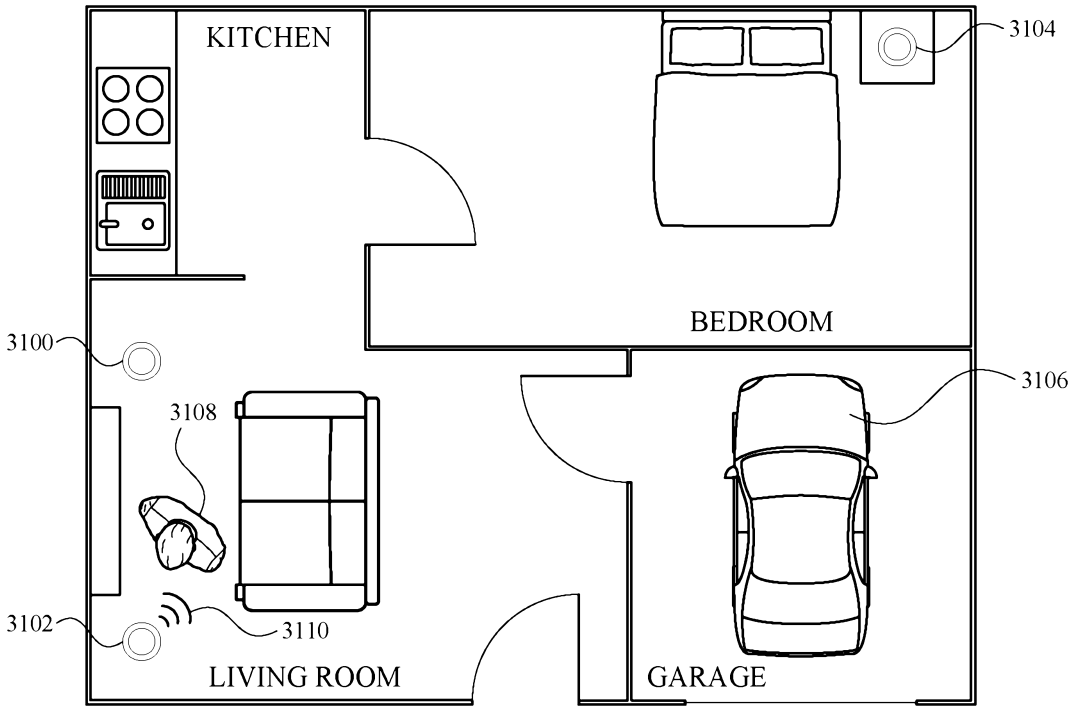

KITCHEN
3104
BEDROOM
3100
3106
3108
3102
3110
LIVING ROOM
GARAGE

FIG. 31D

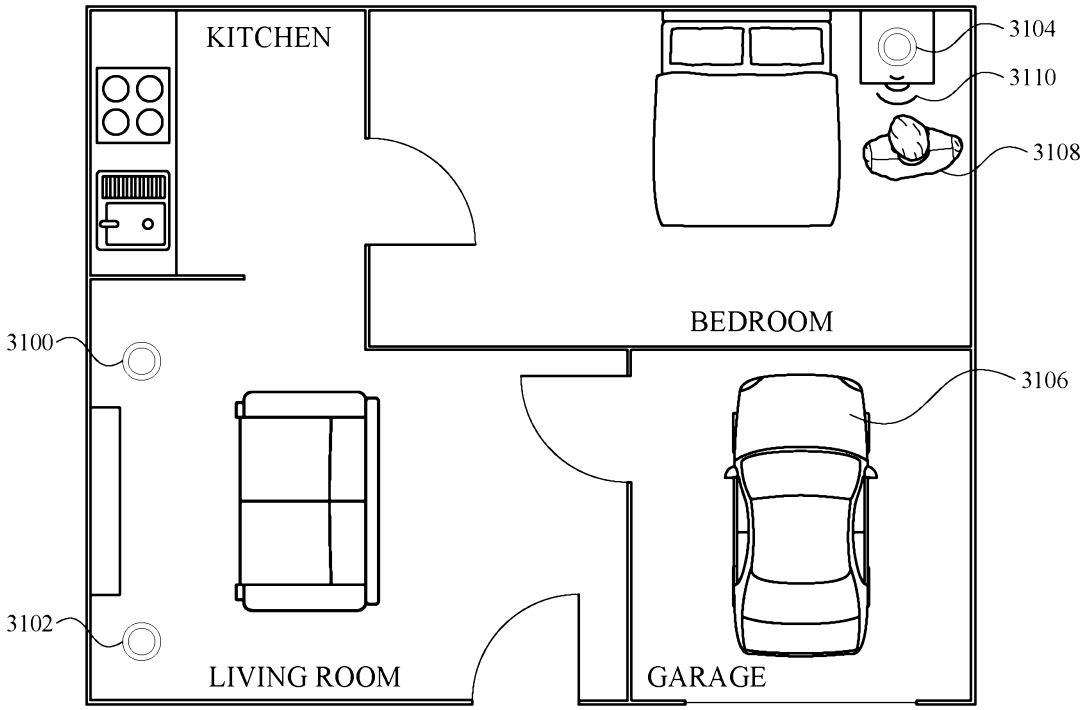

KITCHEN
3104
3110
3108
BEDROOM
3100
3106
3102
LIVING ROOM
GARAGE

FIG. 31E

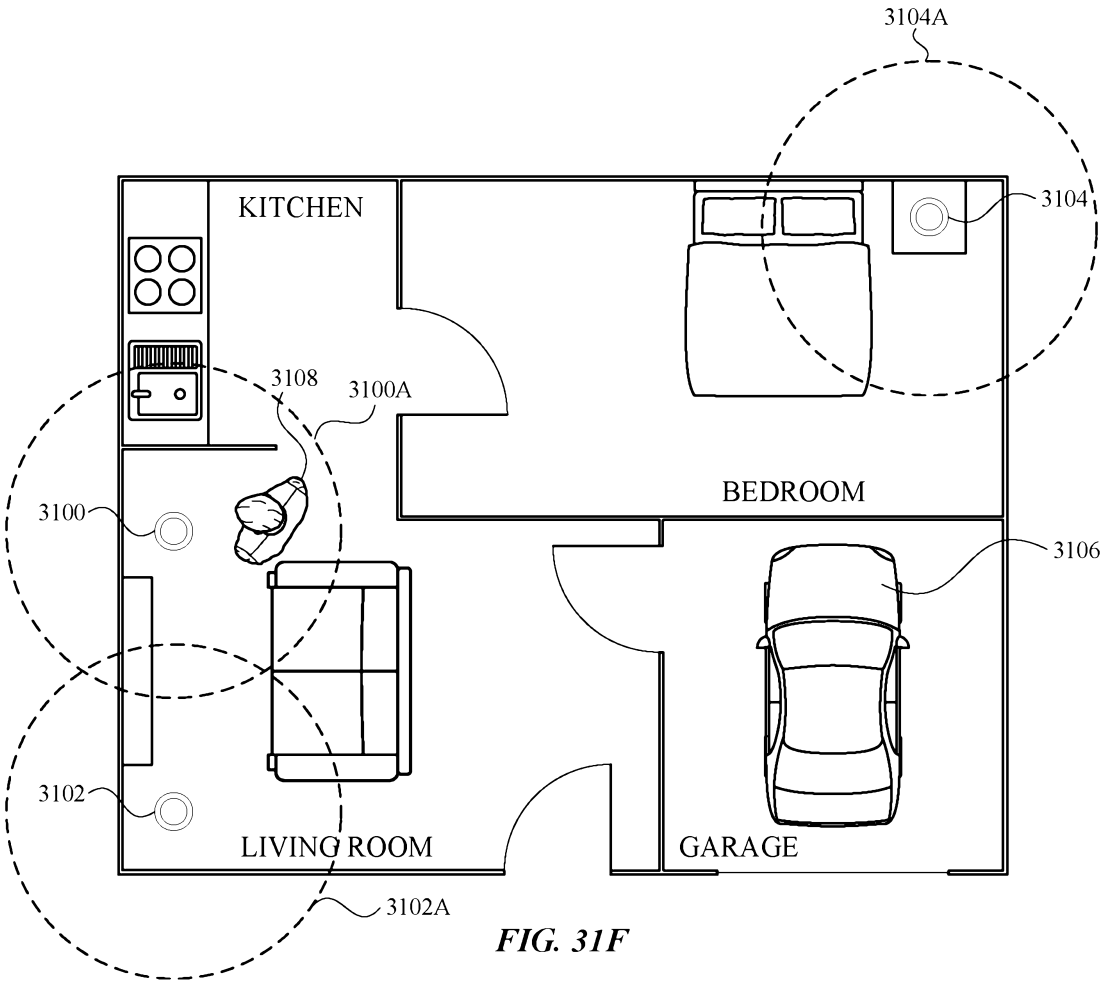

3104A
KITCHEN
3104
3108
3100A
BEDROOM
3100
3106
3102
LIVING ROOM
GARAGE
3102A

FIG. 31F

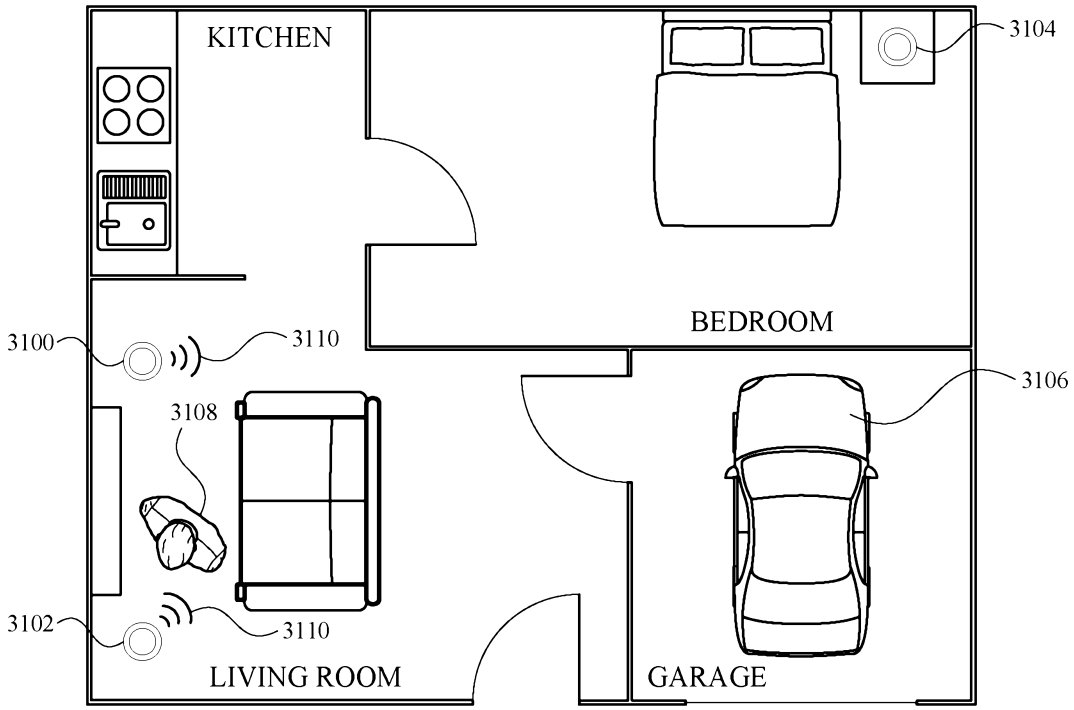

KITCHEN
3104
BEDROOM
3100
3110
3106
3108
3102
3110
LIVING ROOM
GARAGE

FIG. 31G

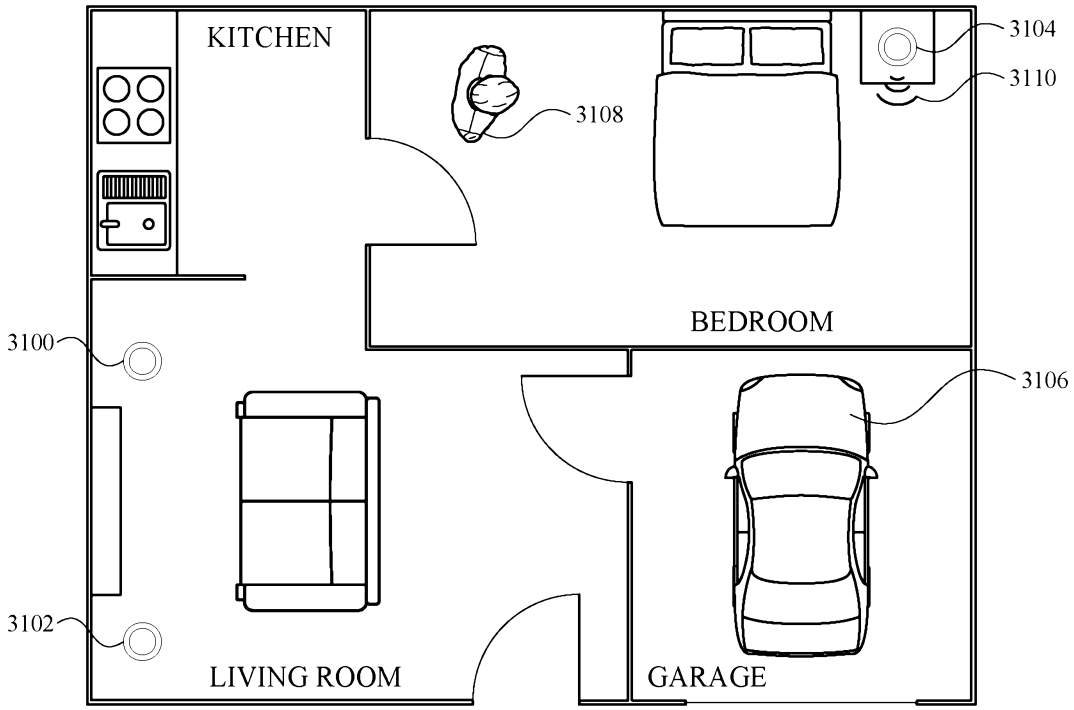

KITCHEN
3104
3110
3108
BEDROOM
3100
3106
3102
LIVING ROOM
GARAGE

FIG. 31H

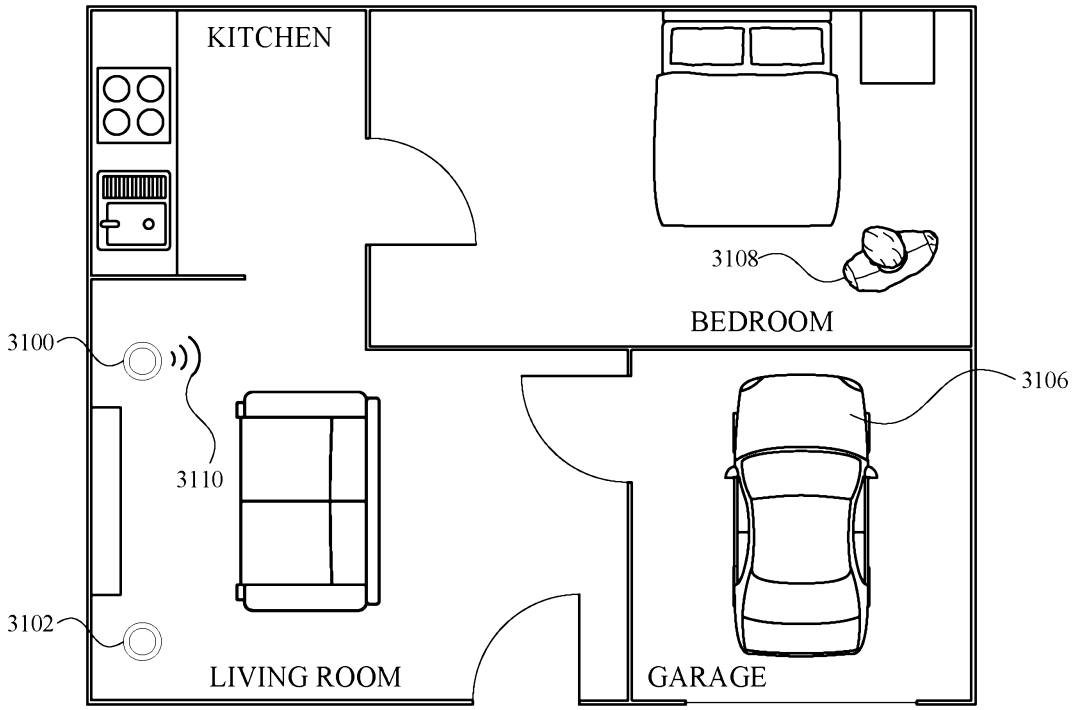

KITCHEN
3108
BEDROOM
3100
3110
3106
3102
LIVING ROOM
GARAGE

FIG. 311

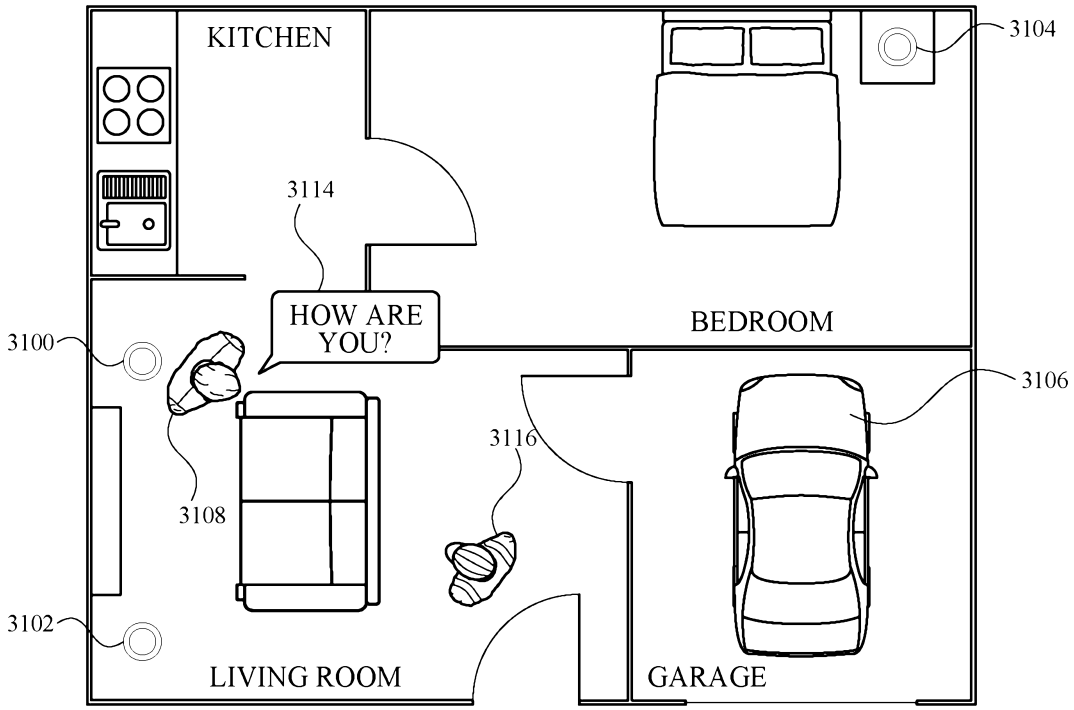

KITCHEN
3104
3114
HOW ARE YOU?
BEDROOM
3100
3106
3116
3108
3102
LIVING ROOM
GARAGE

FIG. 31J

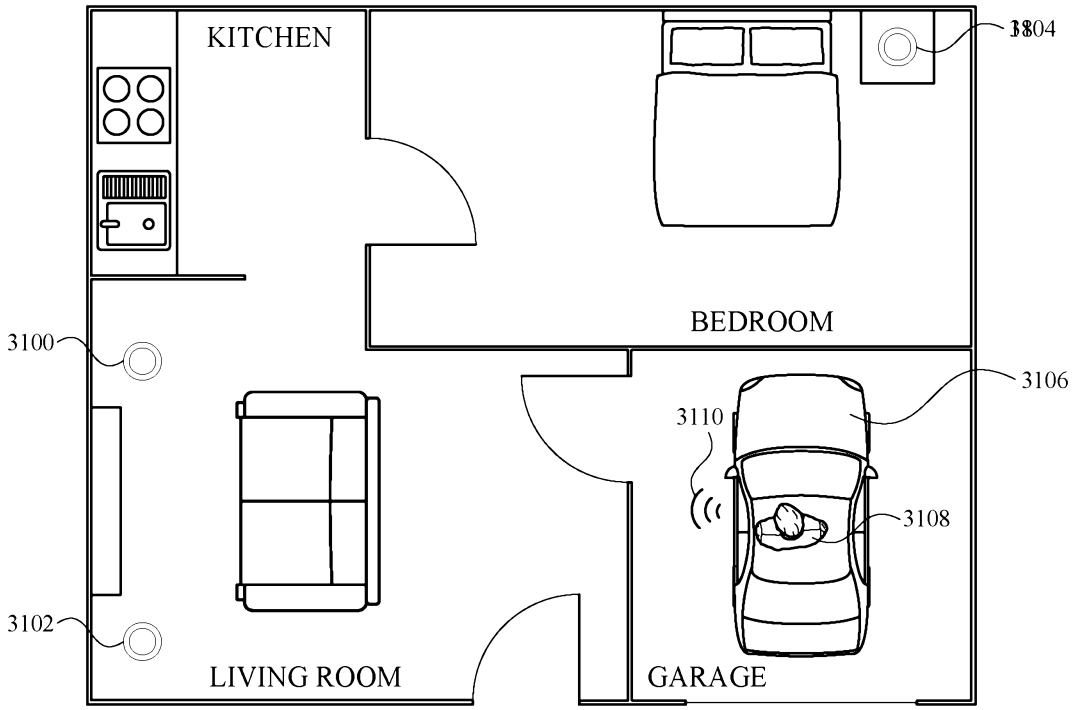

KITCHEN
3104
BEDROOM
3100
3106
3110
3108
3102
LIVING ROOM
GARAGE

FIG. 31K

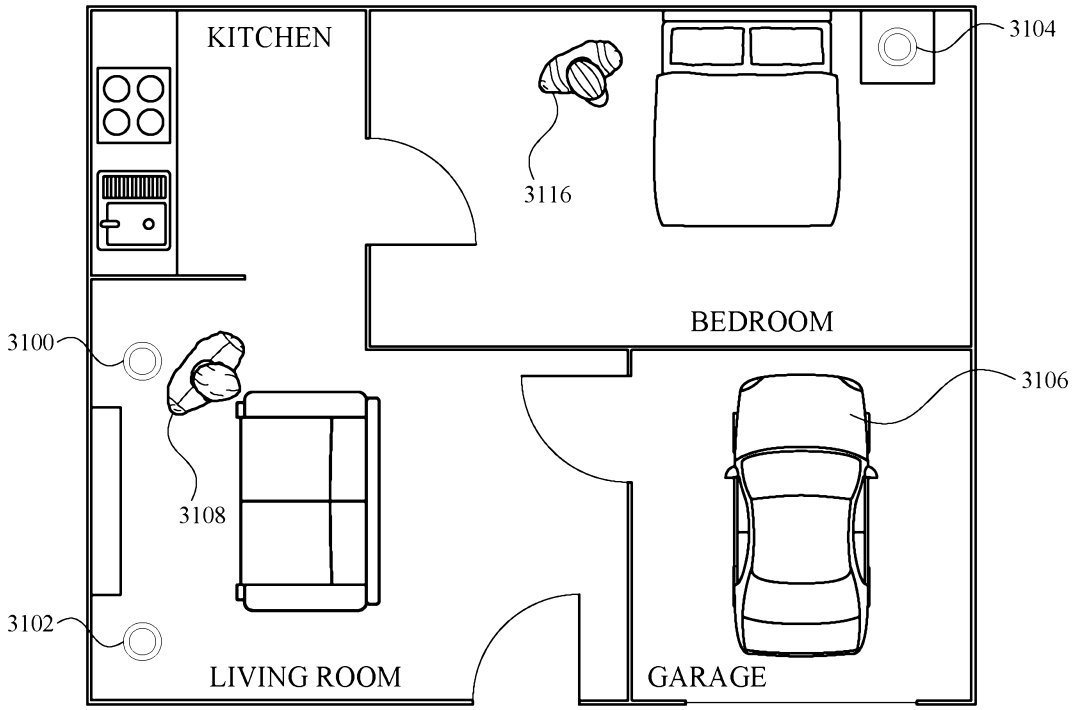

KITCHEN
3104
3116
BEDROOM
3100
3106
3108
3102
LIVING ROOM
GARAGE

FIG. 31L

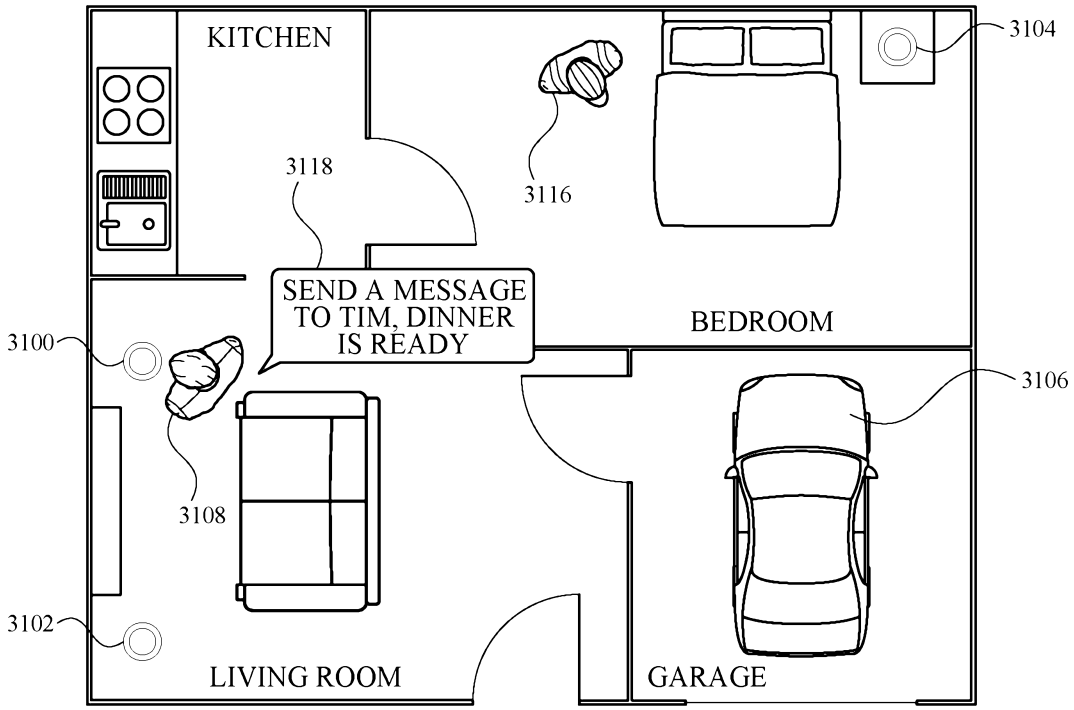

KITCHEN
3104
3118
3116
SEND A MESSAGE TO TIM, DINNER IS READY
BEDROOM
3100
3106
3108
3102
LIVING ROOM
GARAGE

FIG. 31M

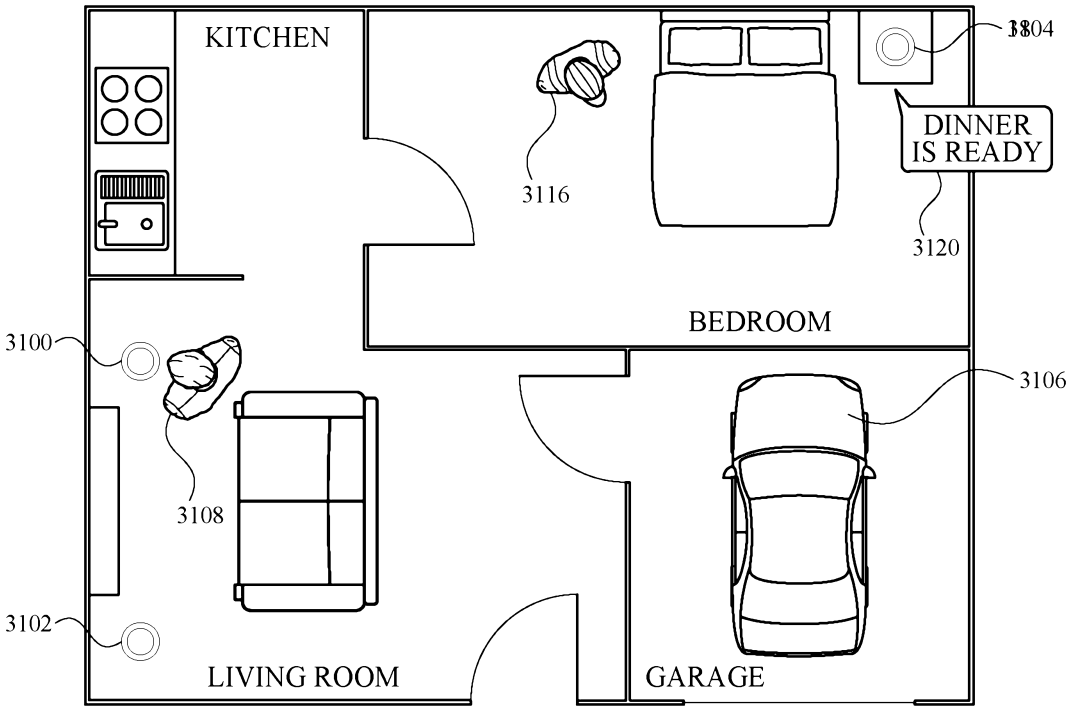

KITCHEN
3804
DINNER
IS READY
3116
3120
BEDROOM
3100
3106
3108
3102
LIVING ROOM
GARAGE

FIG. 31N

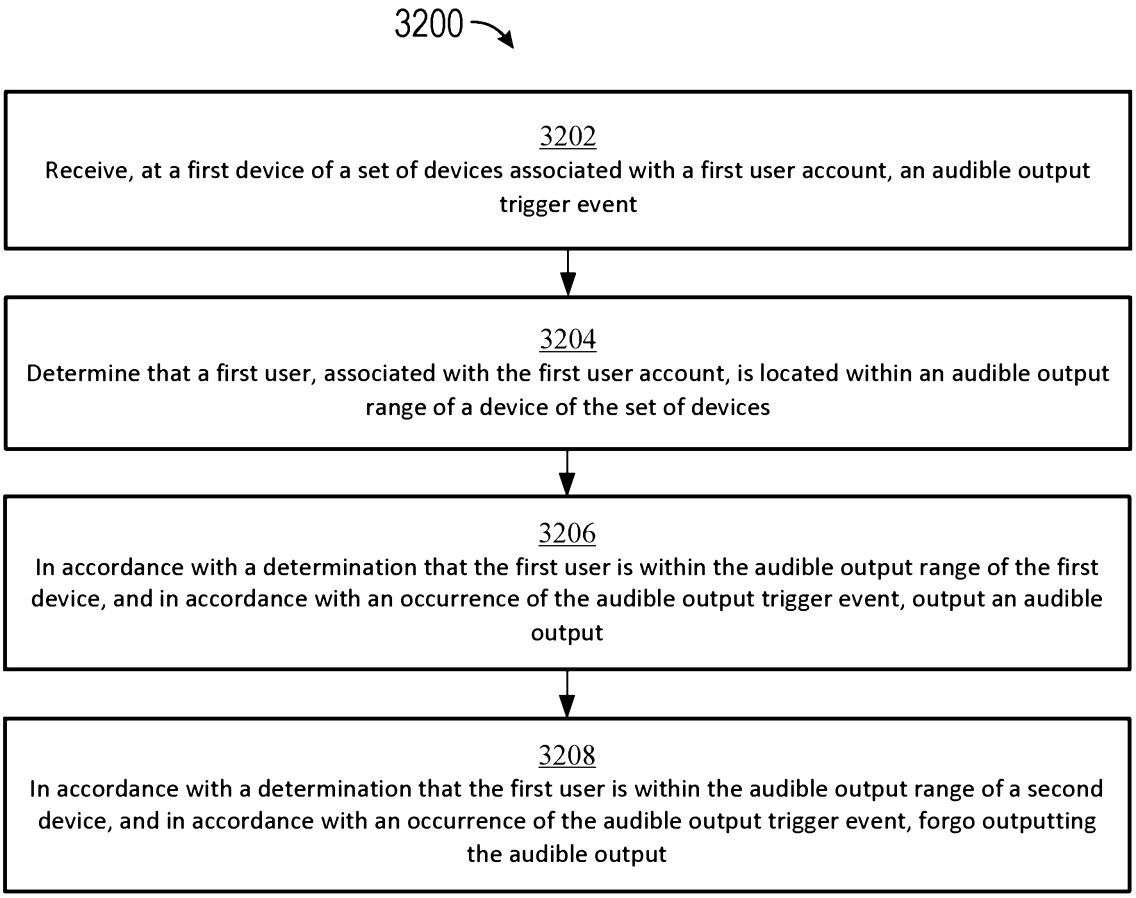

3200
3202
Receive, at a first device of a set of devices associated with a first user account, an audible output trigger event
3204
Determine that a first user, associated with the first user account, is located within an audible output range of a device of the set of devices
3206
In accordance with a determination that the first user is within the audible output range of the first device, and in accordance with an occurrence of the audible output trigger event, output an audible output
3208
In accordance with a determination that the first user is within the audible output range of a second device, and in accordance with an occurrence of the audible output trigger event, forgo outputting the audible output

FIG. 32

WHAT IS CLAIMED IS:

1. A method for providing a digital assistant service, comprising:

at a first electronic device with one or more processors and memory:

receiving, from a first user, a first speech input representing a user request;

obtaining an identity of the first user;

in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device;

receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

providing a representation of the response to the first user.

2. The method of claim 1, wherein the second electronic device is a server remotely disposed from the first electronic device; and the third electronic device is a client device disposed in the vicinity of the first electronic device.

3. The method of any of claims 1 and 2, wherein the third electronic device is a proxy device of a server.

4. The method of any of claims 1-3, wherein the user request comprises a request for information specific to the first user.

5. The method of any of claims 1-4, wherein the user request comprises a request for non-user-specific information.

6. The method of any of claims 1-5, wherein the user request comprises a request for performing a task.

7. The method of any of claims 1-6, wherein obtaining the identity of the first user comprises:

obtaining authentication data associated with the first user; and

obtaining a determination of the identity of the first user based on the authentication data.

8. The method of claim 7, wherein obtaining the determination of the identity of the first user based on the authentication data comprises:

providing the authentication data to at least one of the second electronic device or the third electronic device; and

receiving the identity of the first user from at least one of the second electronic device or the third electronic device, wherein the identity of the first user is determined based on the authentication data.

9. The method of any of claims 1-8, wherein providing the representation of the user request to at least one of the second electronic device or the third electronic device comprises:

determining whether the third electronic device is communicatively coupled to the first electronic device;

in accordance with a determination that the third electronic device is communicatively coupled to the first electronic device, providing the representation of the user request to the third electronic device and not to the second electronic device; and

in accordance with a determination that the third electronic device is not communicatively coupled to the first electronic device, providing the representation of the user request to the second electronic device.

10. The method of any of claims 1-9, wherein providing the representation of the user request to at least one of the second electronic device or the third electronic device comprises:

providing the representation of the user request to the second electronic device and not to the third electronic device.

11. The method of any of claims 1-8, wherein providing the representation of the user request to at least one of the second electronic device or the third electronic device comprises:

providing the representation of the user request to both the second electronic device and the third electronic device.

12. The method of any of claims 1-9, wherein the representation of the user request is provided to the third electronic device and not to the second electronic device; and wherein receiving the response to the user request comprises:

causing the third electronic device to determine whether the third electronic device is capable of providing the response to the user request; and

in accordance with a determination that the third electronic device is capable of providing the response to the user request, receiving, at the first electronic device, the response to the user request from the third electronic device.

13. The method of claim 12, further comprising:

in accordance with a determination that the third electronic device is incapable of providing the response to the user request,

forwarding, by the third electronic device, the representation of the user request to the second electronic device; and

receiving, at the first electronic device, the response to the user request from the second electronic device.

14. The method of any of claims 1-10, wherein the representation of the user request is provided to the second electronic device and not to the third electronic device, and wherein receiving the response to the user request comprises:

causing the second electronic device to determine whether the second electronic device is capable of providing the response to the user request; and

in accordance with a determination that the second electronic device is capable of providing the response to the user request, receiving, at the first electronic device, the response to the user request from the second electronic device.

15. The method of claim 14, further comprising:

in accordance with a determination that the second electronic device is incapable of providing the response to the user request,

forwarding, by the second electronic device, the representation of the user request to the third electronic device; and

receiving, at the first electronic device, the response to the user request.

16. The method of claim 15, wherein the first electronic device receives the response to the user request from the third electronic device.

17. The method of claim 15, wherein the first electronic device receives the response to the user request from the second electronic device based on a response provided by the third electronic device to the second electronic device.

18. The method of any of claims 1-8 and 11, wherein the representation of the user request is provided from the first electronic device to both the second electronic device and the third electronic device, and wherein receiving the response to the user request comprises at least one of:

causing the second electronic device to determine whether the second electronic device is capable of providing the response to the user request; and

causing the third electronic device to determine whether the third electronic device is capable of providing the response to the user request.

19. The method of claim 18, further comprising:

in accordance with a determination that the second electronic device is capable of providing the response to the user request, and that the third electronic device is incapable of providing the response to the user request,

receiving, at the first electronic device, the response to the user request from the second electronic device.

20. The method of claim 18, further comprising:

in accordance with a determination that the third electronic device is capable of providing the response to the user request, and that the second electronic device is incapable of providing the response to the user request,

receiving, at the first electronic device, the response to the user request from the third electronic device.

21. The method of claim 18, further comprising:

in accordance with a determination that both the second electronic device and the third electronic device are capable of providing the response to the user request,

receiving, at the first electronic device, the response to the user request from the second electronic device or the third electronic device based on a predetermined condition.

22. The method of any of claims 1-21, wherein providing the representation of the response to the first user comprises:

providing, at the first electronic device, a speech output including information in response to the user request.

23. The method of claim 22, wherein the information is provided by the second electronic device or the third electronic device.

24. The method of any of claims 1-23, wherein providing the representation of the response to the first user comprises:

providing, at the first electronic device, a speech output associated with performing a task in accordance with the user request.

25. The method of claim 24, wherein the task is performed by the third electronic device.

26. The method of any of claims 24 and 25, wherein the task is performed by the first electronic device and the third electronic device.

27. The method of claim 26, wherein the task is further performed by one or more additional electronic devices.

28. The method of any of claims 1-27, further comprising, prior to providing the representation of the response to the first user:

determining whether the representation of the response is to be provided by the first electronic device;

in accordance with a determination that the representation of the response is to be provided by the first electronic device, providing the representation of the response to the first user by the first electronic device; and

in accordance with a determination that the representation of the response is not to be provided by the first electronic device, forwarding the response to one or more additional electronic devices.

29. The method of claim 28, wherein determining whether the response is to be provided by the first electronic device is based on the user request.

30. The method of any of claims 28 and 29, wherein determining whether the response is to be provided by the first electronic device is based on at least one of detecting a location of the user or tracking the user's movement.

31. The method of any of claims 1-30, further comprising, prior to receiving the first speech input,

receiving a second speech input including a predetermined content; and

in response to receiving the second speech input, activating the first electronic device.

32. The method of claim 31, wherein the second speech input does not cause one or more additional electronic devices to be activated, the one or more additional electronic devices being disposed in the vicinity of the first electronic device.

33. The method of any of claims 31 and 32, wherein obtaining the identity of the first user is based on the second speech input.

34. The method of any of claims 1-33, further comprising, prior to providing the representation of the user request to at least one of the second electronic device and the third electronic device:

outputting one or more speech inquiries regarding the user request; and

receiving an additional speech input from the first user in response to the one or more speech inquiries.

35. The method of any of claims 1-34, further comprising, prior to providing the representation of the user request to at least one of the second electronic device or the third electronic device:

establishing a connection between the first electronic device and at least one of the second electronic device or the third electronic device.

36. The method of claim 35, wherein establishing the connection is based on a near-field communication between the first electronic device and the third electronic device.

37. The method of any of claims 35 and 36, wherein establishing the connection is based on detecting of the third electronic device being within a predetermined distance from the first electronic device.

38. The method of any of claims 35-37, wherein establishing the connection is based on a previous established connection between the first electronic device and the third electronic device.

39. The method of any of claims 35-38, further comprising:

establishing a connection between the first electronic device and the third electronic device, wherein the third electronic device is associated with the first user; and

establishing a connection between the first electronic device and a fourth electronic device, wherein the fourth electronic device is associated with a second user.

40. The method of any of claims 35-39, further comprising:

after establishing a connection between the first electronic device and the third electronic device, notifying the second electronic device of the established connection..

41. The method of any of claims 1-39, further comprising, establishing one or more connections between the first electronic device and one or more additional electronic devices, wherein the additional electronic devices are the same type of devices as the first electronic device.

42. The method of any of claims 1-41, further comprising, while providing the response to the first user by the first electronic device,

determining whether the response is to be continually provided at a different electronic device; and

in accordance with a determination that the response is to be continually provided at a different electronic device, causing the response to be continually provided by at least one of the third electronic device or one or more additional electronic devices.

43. The method of claim 42, wherein the determination of whether the response is to be continually provided at a different electronic device is based on a third speech input.

44. The method of any of claims 42 and 43, wherein the determination of whether the response is to be continually provided at a different electronic device is based on detecting whether the first user's location variation with respect to the first electronic device satisfies a predetermined condition.

45. The method of any of claims 1-44, further comprising, after providing the response to the first user, monitoring subsequent speech inputs.

46. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors of a first electronic device, cause the first electronic device to:

receive, from a first user, a first speech input representing a user request;

obtain an identity of the first user;

in accordance with the user identity, provide a representation of the user request to at least one of a second electronic device or a third electronic device;

receive, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

provide a representation of the response to the first user.

47. A first electronic device, comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for:

receiving, from a first user, a first speech input representing a user request;

obtaining an identity of the first user;

in accordance with the user identity, providing a representation of the user request to at least one of a second electronic device or a third electronic device;

receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

providing a representation of the response to the first user.

48. A first electronic device, comprising:

means for receiving, from a first user, a first speech input representing a user request;

means for obtaining an identity of the first user;

in accordance with the user identity, means for providing a representation of the user request to at least one of a second electronic device or a third electronic device;

means for receiving, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide a response to the first electronic device, the response to the user request from the second electronic device or the third electronic device; and

means for providing a representation of the response to the first user.

49. A first electronic device comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for performing the method of any one of claims 1-45.

50. A first electronic device, comprising:

means for performing any of the methods of claims 1-45.

51. A non-transitory computer-readable storage medium comprising one or more programs for execution by one or more processors of a first electronic device, the one or more programs including instructions which, when executed by the one or more processors, cause the first electronic device to perform the method of any one of claims 1-45.

52. A system for operating a digital assistant, the system comprising means for performing any of the methods of claims 1-45.

53. A method for providing a digital assistant service, comprising:

at a first electronic device with one or more processors and memory:

receiving a notification of an event associated with a first user;

in response to receiving the notification, outputting an indication of the notification;

receiving one or more speech inputs;

in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

54. The method of claim 53, wherein the notification of the event includes a representation of at least one of an incoming call, a reminder, a message, a voicemail, or a news alert.

55. The method of any of claims 53 and 54, wherein receiving the notification of the event comprises:

receiving the notification from at least one of a second electronic device or a third electronic device.

56. The method of any of claims 53-55, wherein outputting an indication of the notification comprises:

outputting the indication of the notification by the first electronic device or one of the additional electronic devices communicatively coupled to the first electronic device.

57. The method of any of claims 53-56, wherein receiving the one or more speech inputs comprises:

receiving a first speech input regarding the notification;

outputting a response in accordance with the notification of the event; and

receiving a second speech input.

58. The method of any of claims 53-57, wherein determining whether the notification is to be provided at the first electronic device comprises:

obtaining an identity of the user who provides at least one of the one or more speech inputs;

determining, based on the identity of the user who provides at least one of the one or more speech inputs and based on the notification, whether the notification is to be provided to the user who provides at least one of the one or more speech inputs; and

in accordance with a determination that the notification is to be provided to the user who provides at least one of the one or more speech inputs, determining whether the notification is to be provided at the first electronic device.

59. The method of claim 58, wherein obtaining the identity of the user who provides at least one of the one or more speech inputs comprises:

obtaining authentication data associated with the user who provides at least one of the one or more speech inputs; and

obtaining a determination of the identity of the user who provides at least one of the one or more speech inputs based on the authentication data.

60. The method of claim 59, wherein obtaining the determination of the identity of the user who provides at least one of the one or more speech inputs comprises:

providing the authentication data to at least one of a second electronic device and a third electronic device; and

receiving the identity of the user who provides at least one of the one or more speech inputs from at least one of the second electronic device and the third electronic device, wherein the identity of the user who provides at least one of the one or more speech inputs is determined based on the authentication data.

61. The method of any of claims 53-60, wherein providing the notification at the first electronic device comprises:

providing an audio output associated with the notification at the first electronic device.

62. The method of any of claims 53-61, further comprising:

in accordance with a determination that the notification is not to be provided at the first electronic device, determining an additional electronic device for providing the notification.

63. The method of claim 62, wherein determining the additional electronic device for providing the notification is based on the one or more speech inputs.

64. The method of any of claims 62 and 63, wherein determining the additional electronic device for providing the notification is based on context information.

65. The method of any of claims 53-64, further comprising, prior to receiving the notification, establishing a connection between the first electronic device and at least one of a second electronic device or a third electronic device.

66. The method of claim 65, wherein establishing the connection is based on a near-field communication between the first electronic device and the third electronic device.

67. The method of any of claims 65 and 66, wherein establishing the connection is based on detecting of the third electronic device being within a predetermined distance from the first electronic device.

68. The method of any of claims 65-67, wherein establishing the connection is based on a previous established connection between the first electronic device and the third electronic device.

69. The method of any of claims 65-68, further comprising:

establishing a connection between the first electronic device and the third electronic device, wherein the third electronic device is associated with the first user; and

establishing a connection between the first electronic device and a fourth electronic device, wherein the fourth electronic device is associated with a second user.

70. The method of any of claims 65-69, further comprising: after establishing a connection between the first electronic device and the third electronic device, notifying the second electronic device of the established connection.

71. The method of any of claims 53-70, further comprising, establishing one or more connections between the first electronic device and one or more additional electronic devices, wherein the one or more additional electronic devices are the same type of devices as the first electronic device.

72. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors of a first electronic device, cause the first electronic device to:

receive a notification of an event associated with a first user;

in response to receiving the notification, output an indication of the notification;

receive one or more speech inputs;

in accordance with the one or more speech inputs, determine whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

73. A first electronic device, comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for:

receiving a notification of an event associated with a first user;

in response to receiving the notification, outputting an indication of the notification;

receiving one or more speech inputs;

in accordance with the one or more speech inputs, determining whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, providing the notification at the first electronic device.

74. An electronic device, comprising:

means for receiving a notification of an event associated with a first user;

in response to receiving the notification, means for outputting an indication of the notification;

means for receiving one or more speech inputs;

in accordance with the one or more speech inputs, means for determining whether the notification is to be provided at the first electronic device; and

in accordance with a determination that the notification is to be provided at the first electronic device, means for providing the notification at the first electronic device.

75. A first electronic device comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for performing the method of any one of claims 53-71.

76. A first electronic device, comprising:

means for performing any of the methods of claims 53-71.

77. A non-transitory computer-readable storage medium comprising one or more programs for execution by one or more processors of a first electronic device, the one or more programs including instructions which, when executed by the one or more processors, cause the first electronic device to perform the method of any one of claims 53-71.

78. A system for operating a digital assistant, the system comprising means for performing any of the methods of claims 53-71.

79. A method for providing a digital assistant service, comprising:

at a first electronic device with one or more processors and memory:

receiving, from a first user, a first speech input representing a user request;

obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

causing the second electronic device to provide at least a portion of the response to the user request.

80. The method of claim 79, wherein obtaining the capability data comprises:

obtaining device capabilities associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device.

81. The method of claim 80, wherein the device capabilities comprise one or more physical attributes associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device.

82. The method of any of claims 80 and 81, wherein the device capabilities comprise data that are capable of being provided by the one of more electronic devices capable of being communicatively coupled to the first electronic device.

83. The method of any of claims 79-82, wherein the first electronic device and the electronic devices capable of being communicatively coupled to the first electronic device are disposed within a predetermined boundary.

84. The method of any of claims 79-83, wherein identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the first user comprises:

obtaining, based on the first speech input, one or more steps for responding to the user request; and

identifying, based on the capability data, the second electronic device for performing at least one step for responding to the user request.

85. The method of claim 84, wherein obtaining one or more steps for responding to the user request comprises at least one of:

receiving, from a third electronic device remotely located from the first electronic device, a plan for responding to the user request; and

determining, by the first electronic device, a plan for responding to the user request, wherein the plan comprising one or more steps for responding to the user request.

86. The method of any of claims 84 and 85, further comprising:

identifying, based on the capability data, one or more additional electronic devices for performing the remaining steps for responding the user request.

87. The method of any of claims 79-86, wherein causing the second electronic device to provide at least a portion of the response to the user request comprises:

causing the second electronic device to perform at least one step for responding to the user request.

88. The method of claim 87, further comprising:

causing one or more additional electronic devices to perform the remaining steps for responding to the user request.

89. The method of any of claims 79-88, further comprising, prior to causing the second electronic device to provide at least a portion of the response to the user request:

providing a first speech output to the first user regarding providing at least a portion of the response by the second electronic device;

receiving a second speech input from the first user; and

in response to receiving the second speech input, determining whether at least a portion of the response is to be provided by the second electronic device.

90. The method of claim 89, wherein providing a first speech output to the first user regarding providing at least a portion of the response at the second electronic device comprises:

annotating one or more steps for responding to the user request; and

providing the first speech output to the first user based on the annotation of one or more steps.

91. The method of any of claims 79-90, further comprising, prior to obtaining capability data associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device:

establishing a connection between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

92. The method of claim 91, wherein establishing the connection is based on a near-field communication between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

93. The method of any of claims 91 and 92, wherein establishing the connection is based on detecting of the one or more electronic devices capable of being communicatively coupled to the first electronic device being within a predetermined distance from the first electronic device.

94. The method of any of claims 91-93, wherein establishing the connection is based on one or more previous established connections between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device.

95. The method of any of claims 79-94, further comprising, prior to obtaining capability data associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device:

inquiring the first user regarding accessing, by the first electronic device, the one or more electronic devices capable of being communicatively coupled to the first electronic device;

receiving a third speech input from the first user; and

in response to receiving the third speech input, determining whether the first electronic device is authorized to access the one or more electronic devices capable of being communicatively coupled to the first electronic device.

96. The method of claim 95, further comprising:

providing one or more duration options for accessing the one or more electronic devices capable of being communicatively coupled to the first electronic device;

receiving a selection of a duration option from the first user; and

accessing the one or more electronic devices capable of being communicatively coupled to the first electronic device based on the selected duration option.

97. The method of any of claims 79-96, further comprising, prior to causing the second electronic device to provide at least a portion of the response to the first user:

obtaining an identity of the first user ; and

determining, based on the identity of the first user, whether the first user is authorized to receive at least a portion of the response to the user request.

98. The method of any of claims 97, wherein obtaining the identity of the first user comprises obtaining the identity based on a voice profile.

99. A non-transitory computer-readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by one or more processors of a first electronic device, cause the first electronic device to:

receive, from a first user, a first speech input representing a user request;

obtain capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

in accordance with the capability data, identify, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

cause the second electronic device to provide at least a portion of the response to the first user.

100. A first electronic device, comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for:

receiving, from a first user, a first speech input representing a user request;

obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

causing the second electronic device to provide at least a portion of the response to the first user.

101. An electronic device, comprising:

means for receiving, from a first user, a first speech input representing a user request;

means for obtaining capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device;

means for in accordance with the capability data, identifying, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request; and

means for causing the second electronic device to provide at least a portion of the response to the first user.

102. A first electronic device comprising:

one or more processors;

memory; and

one or more programs stored in memory, the one or more programs including instructions for performing the method of any one of claims 79-98.

103. A first electronic device, comprising:

means for performing any of the methods of claims 79-98.

104. A non-transitory computer-readable storage medium comprising one or more programs for execution by one or more processors of a first electronic device, the one or more programs including instructions which, when executed by the one or more processors, cause the first electronic device to perform the method of any one of claims 79-98.

105. A system for operating a digital assistant, the system comprising means for performing any of the methods of claims 79-98.

DK 2017 70434 A1

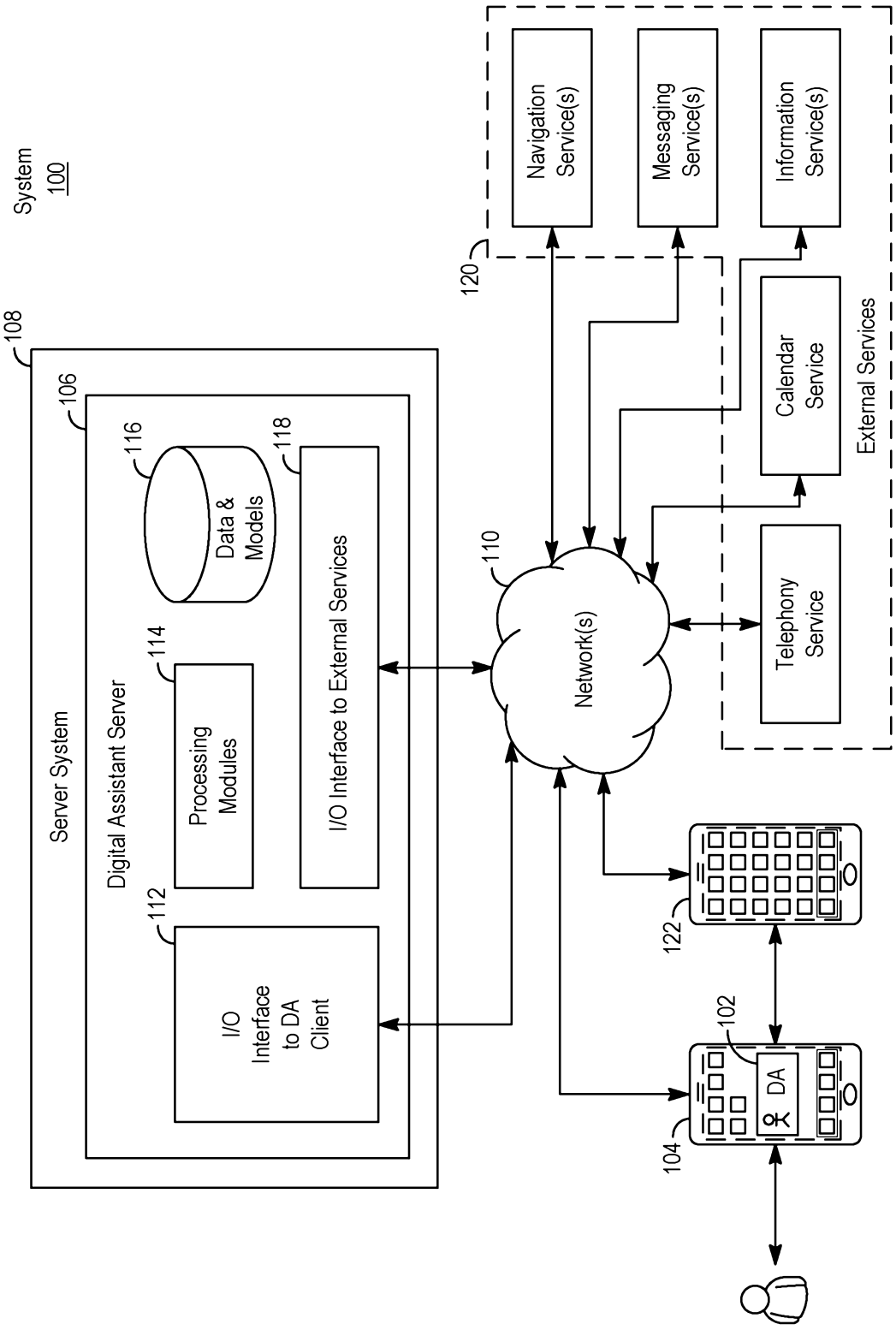


FIG. 1

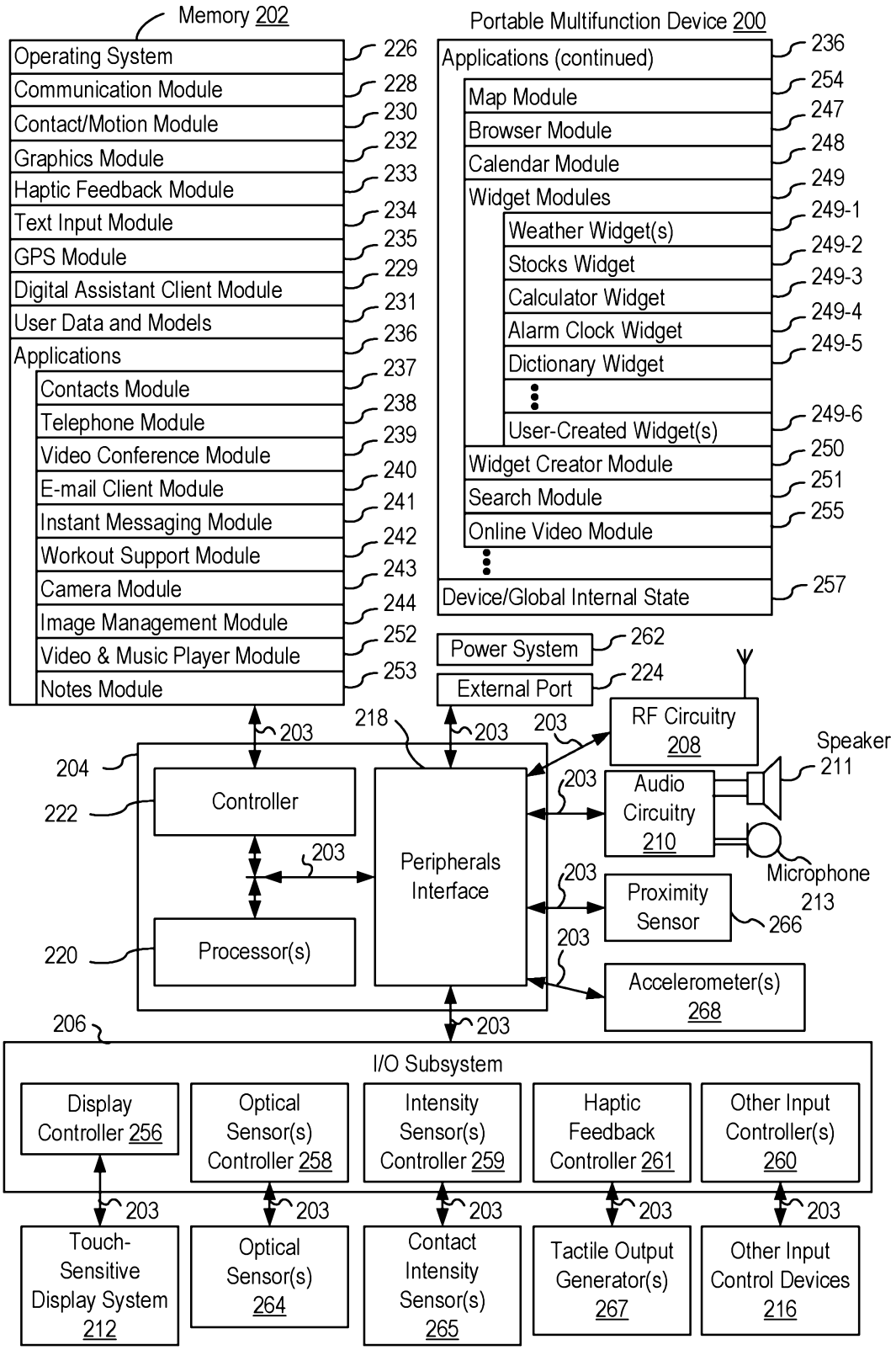

Memory 202
Operating System
Communication Module
Contact/Motion Module
Graphics Module
Haptic Feedback Module
Text Input Module
GPS Module
Digital Assistant Client Module
User Data and Models
Applications
Contacts Module
Telephone Module
Video Conference Module
E-mail Client Module
Instant Messaging Module
Workout Support Module
Camera Module
Image Management Module
Video & Music Player Module
Notes Module
226
228
230
232
233
234
235
229
231
236
237
238
239
240
241
242
243
244
252
253
Portable Multifunction Device 200
Applications (continued)
Map Module
Browser Module
Calendar Module
Widget Modules
Weather Widget(s)
Stocks Widget
Calculator Widget
Alarm Clock Widget
Dictionary Widget
User-Created Widget(s)
Widget Creator Module
Search Module
Online Video Module
Device/Global Internal State
236
254
247
248
249
249-1
249-2
249-3
249-4
249-5
249-6
250
251
255
257
Power System
External Port
262
224
RF Circuitry 208
Speaker 211
Audio Circuitry 210
Microphone 213
Proximity Sensor
266
Accelerometer(s) 268
Controller
Peripherals Interface
Processor(s)
203
204
222
218
220
206
I/O Subsystem
Display Controller 256
Optical Sensor(s) Controller 258
Intensity Sensor(s) Controller 259
Haptic Feedback Controller 261
Other Input Controller(s) 260
Touch-Sensitive Display System 212
Optical Sensor(s) 264
Contact Intensity Sensor(s) 265
Tactile Output Generator(s) 267
Other Input Control Devices 216

FIG. 2A

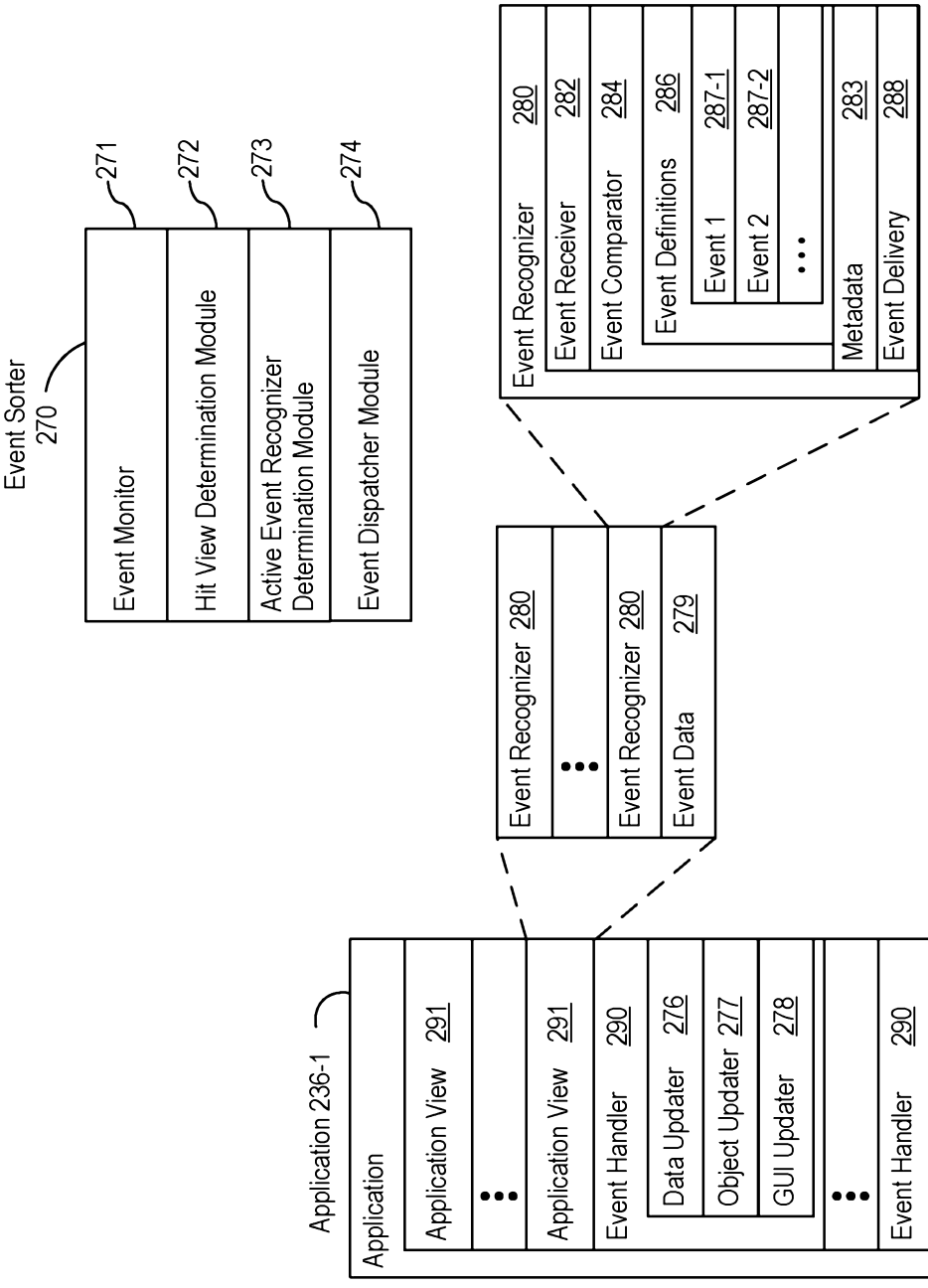

Event Sorter
270
271
Event Monitor
272
Hit View Determination Module
273
Active Event Recognizer Determination Module
274
Event Dispatcher Module
Event Recognizer 280
Event Receiver 282
Event Comparator 284
Event Definitions 286
Event 1 287-1
Event 2 287-2
...
Metadata 283
Event Delivery 288
Event Recognizer 280
...
Event Recognizer 280
Event Data 279
Application 236-1
Application
Application View 291
...
Application View 291
Event Handler 290
Data Updater 276
Object Updater 277
GUI Updater 278
...
Event Handler 290
...
Application Internal State 292

FIG. 2B

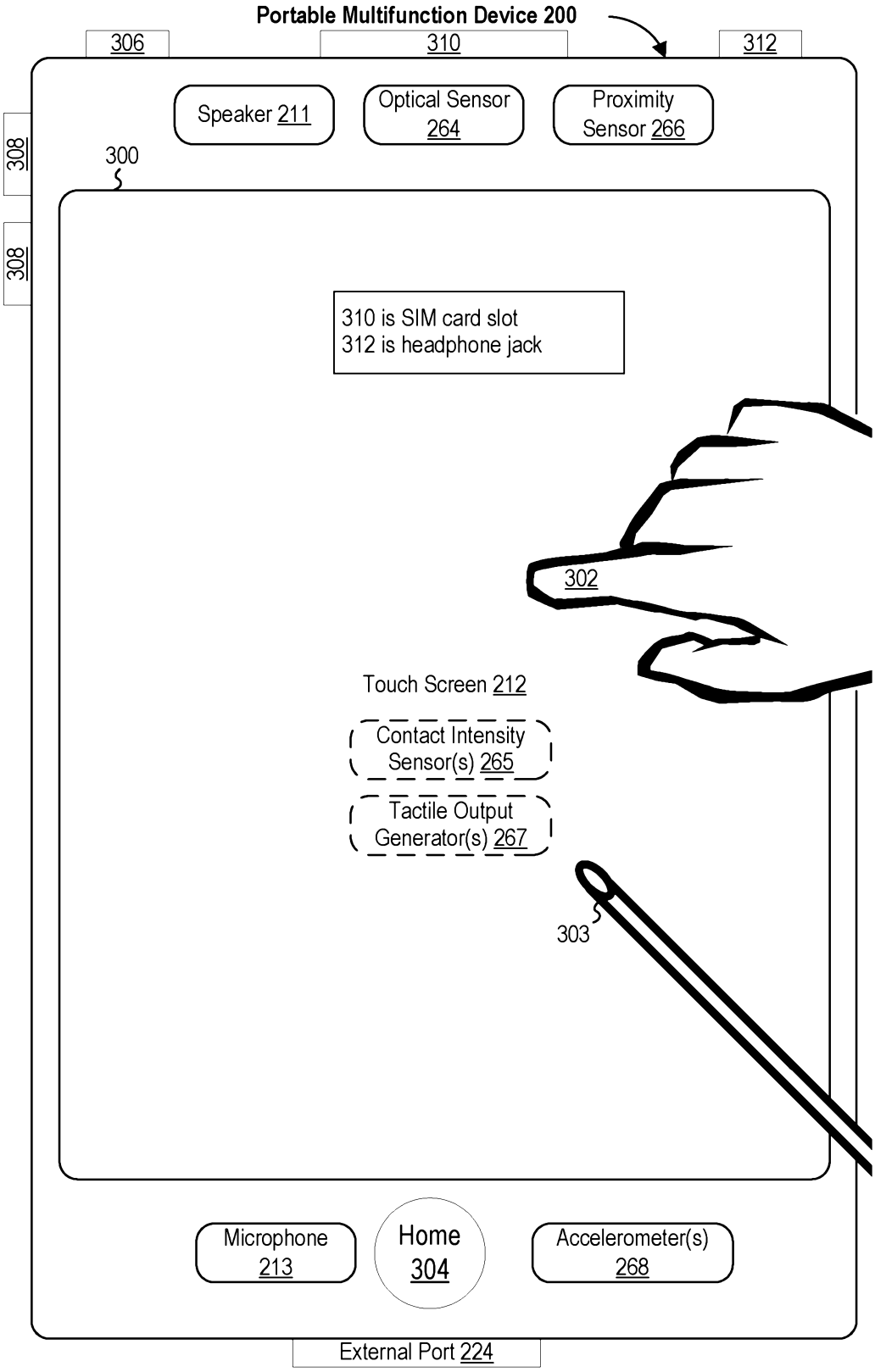

Portable Multifunction Device 200
306
310
312
Speaker 211
Optical Sensor 264
Proximity Sensor 266
308
300
308
310 is SIM card slot
312 is headphone jack
302
Touch Screen 212
Contact Intensity Sensor(s) 265
Tactile Output Generator(s) 267
303
Microphone 213
Home 304
Accelerometer(s) 268
External Port 224

FIG. 3

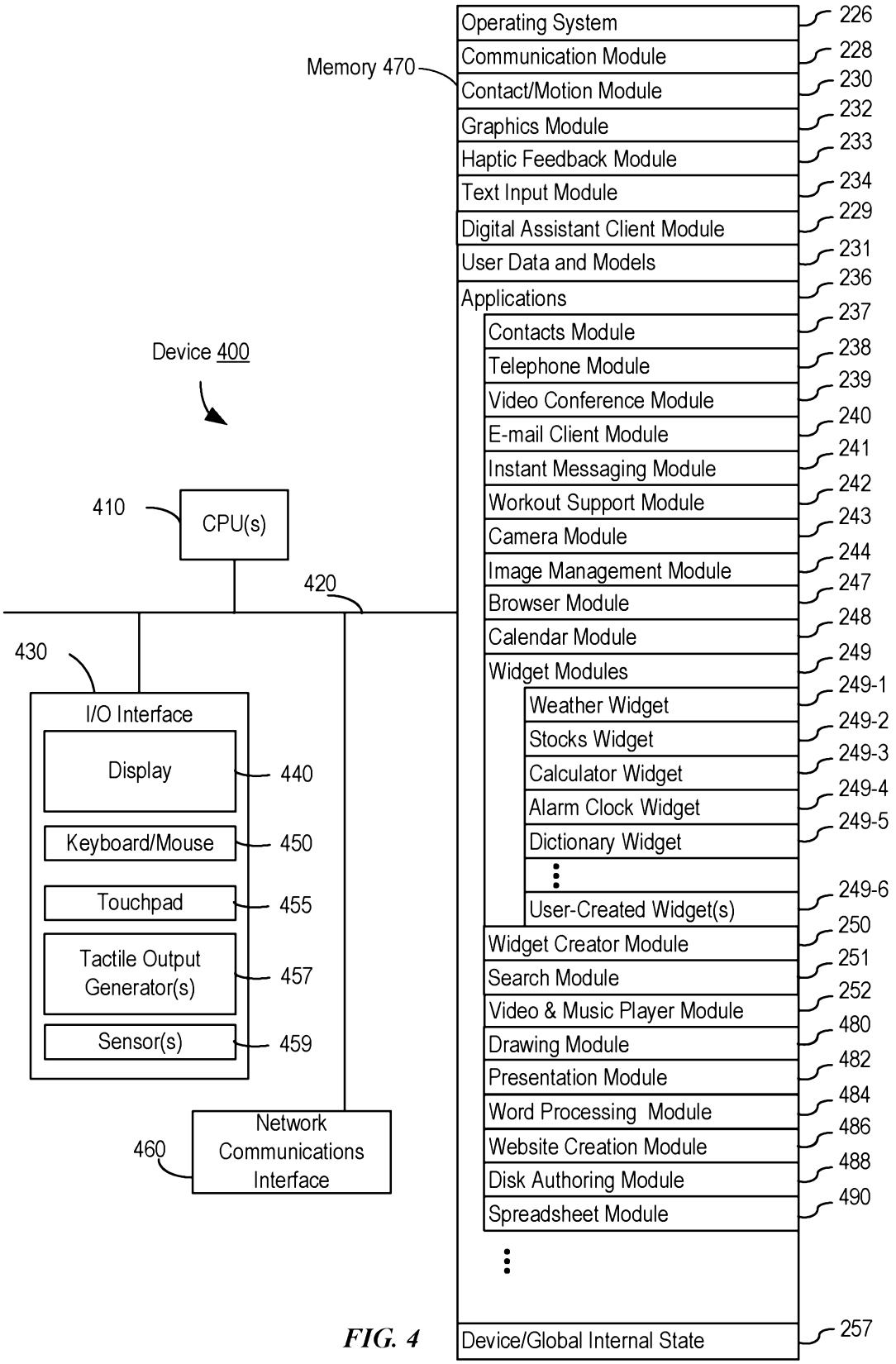

Device 400
410
CPU(s)
420
430
I/O Interface
Display
440
Keyboard/Mouse
450
Touchpad
455
Tactile Output Generator(s)
457
Sensor(s)
459
460
Network Communications Interface
Memory 470
Operating System
226
Communication Module
228
Contact/Motion Module
230
Graphics Module
232
Haptic Feedback Module
233
Text Input Module
234
Digital Assistant Client Module
229
User Data and Models
231
Applications
236
Contacts Module
237
Telephone Module
238
Video Conference Module
239
E-mail Client Module
240
Instant Messaging Module
241
Workout Support Module
242
Camera Module
243
Image Management Module
244
Browser Module
247
Calendar Module
248
Widget Modules
249
Weather Widget
249-1
Stocks Widget
249-2
Calculator Widget
249-3
Alarm Clock Widget
249-4
Dictionary Widget
249-5
User-Created Widget(s)
249-6
Widget Creator Module
250
Search Module
251
Video & Music Player Module
252
Drawing Module
480
Presentation Module
482
Word Processing Module
484
Website Creation Module
486
Disk Authoring Module
488
Spreadsheet Module
490
Device/Global Internal State
257

FIG. 4

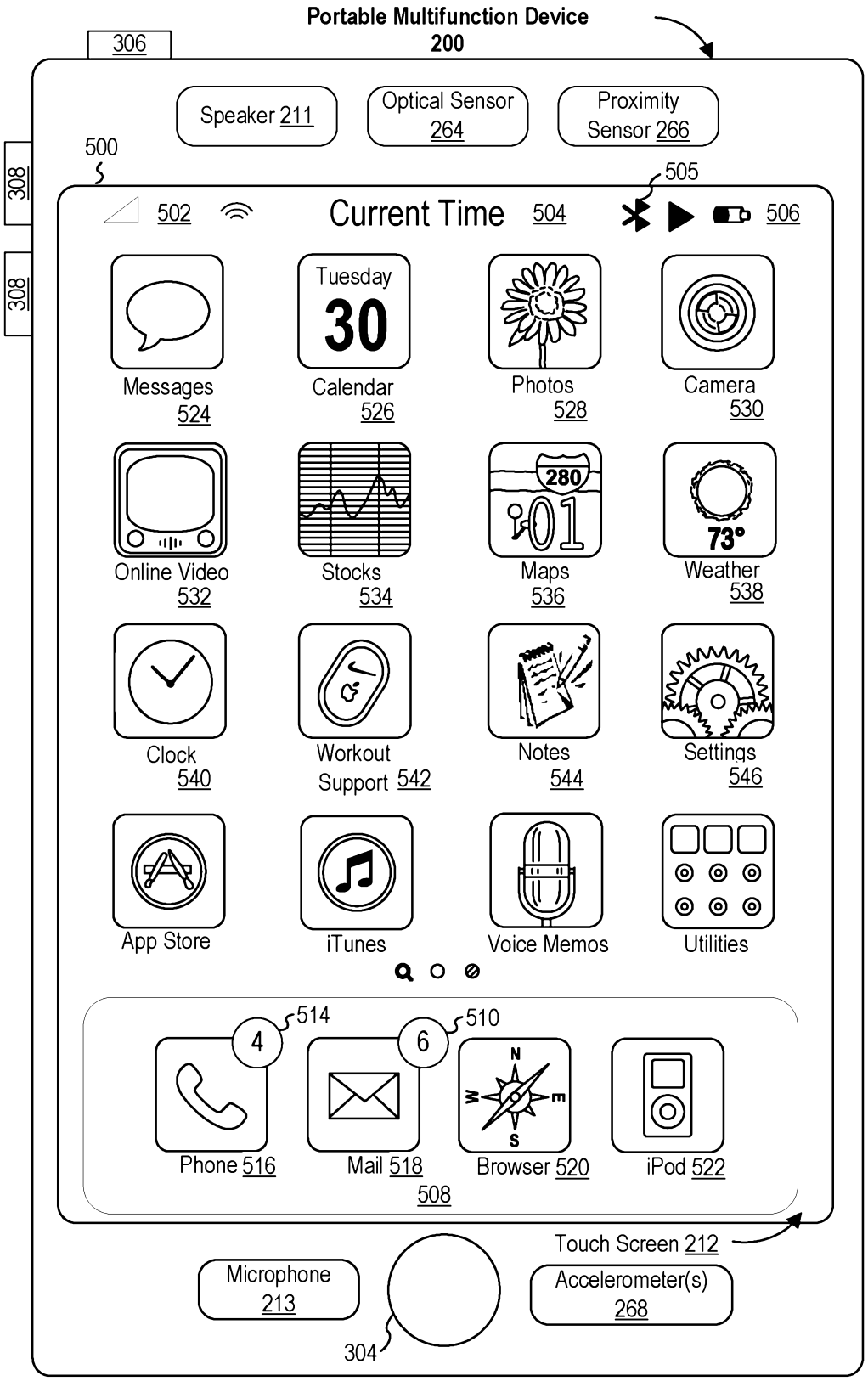
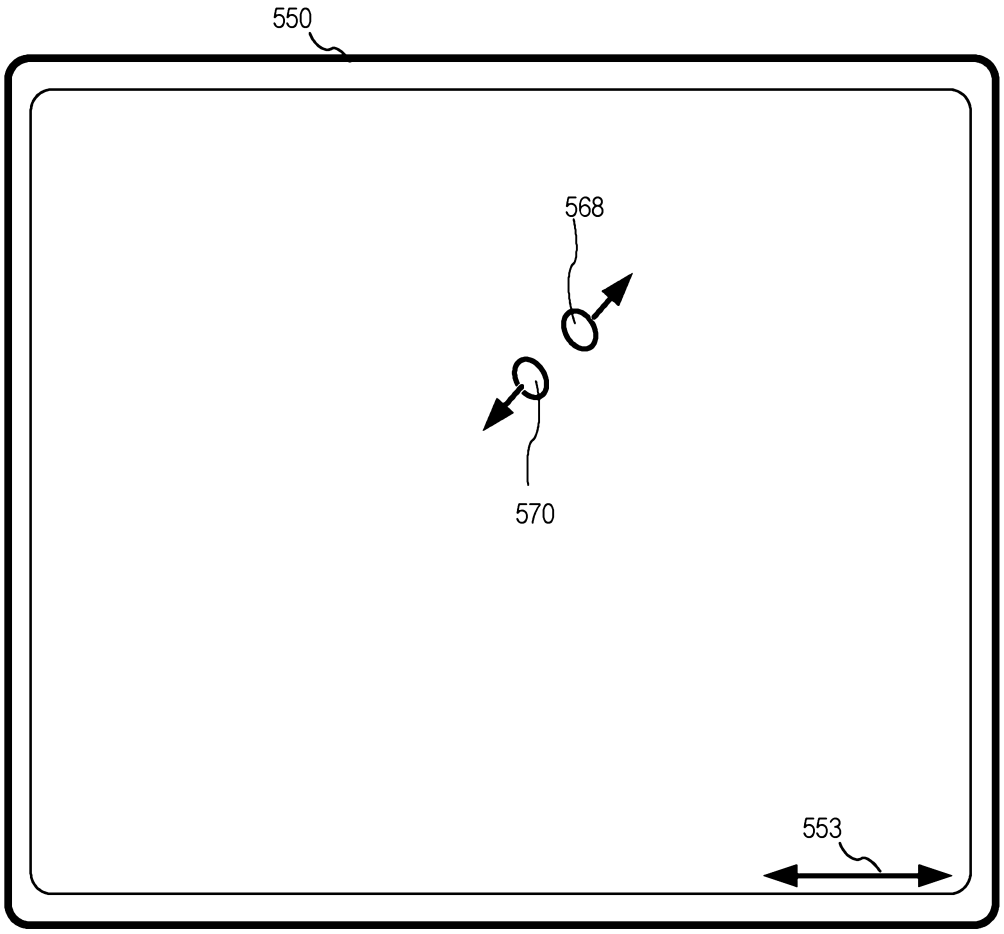

Portable Multifunction Device
200
306
Speaker 211
Optical Sensor 264
Proximity Sensor 266
500
308
308
505
502
Current Time 504
506
Messages 524
Tuesday
30
Calendar 526
Photos 528
Camera 530
Online Video 532
Stocks 534
280
Maps 536
73°
Weather 538
Clock 540
Workout Support 542
Notes 544
Settings 546
App Store
iTunes
Voice Memos
Utilities
514
510
4
6
N
W
E
S
Phone 516
Mail 518
Browser 520
iPod 522
508
Touch Screen 212
Microphone 213
Accelerometer(s) 268
304

FIG. 5A


550
568
570
553

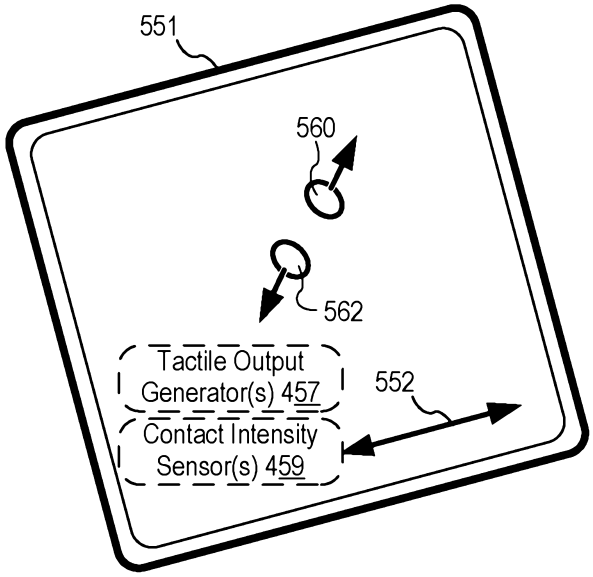

551
560
562
Tactile Output Generator(s) 457
Contact Intensity Sensor(s) 459
552

FIG. 5B

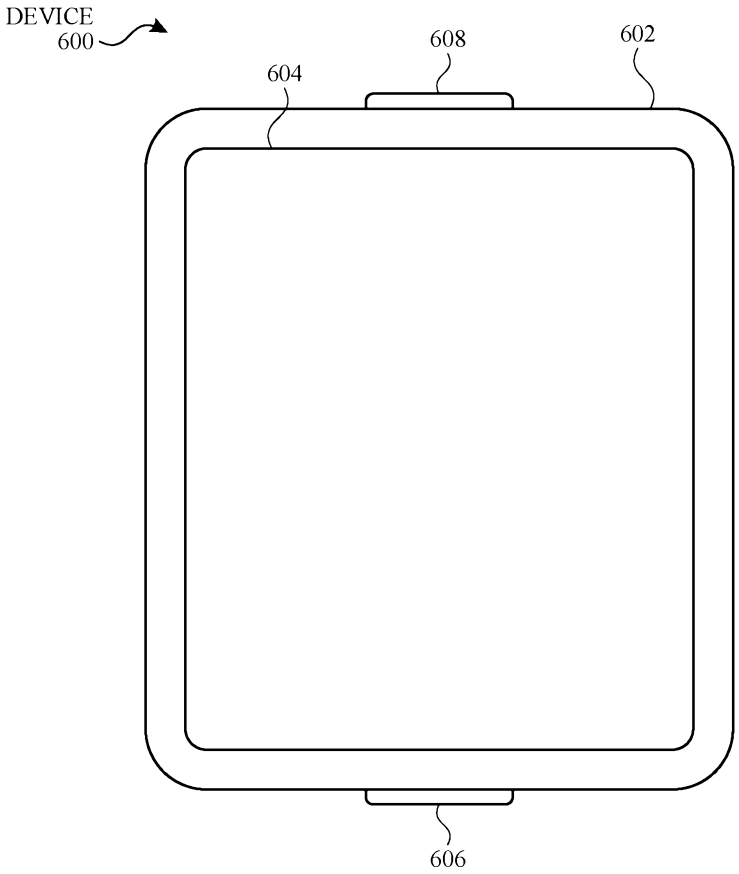


FIG. 6A

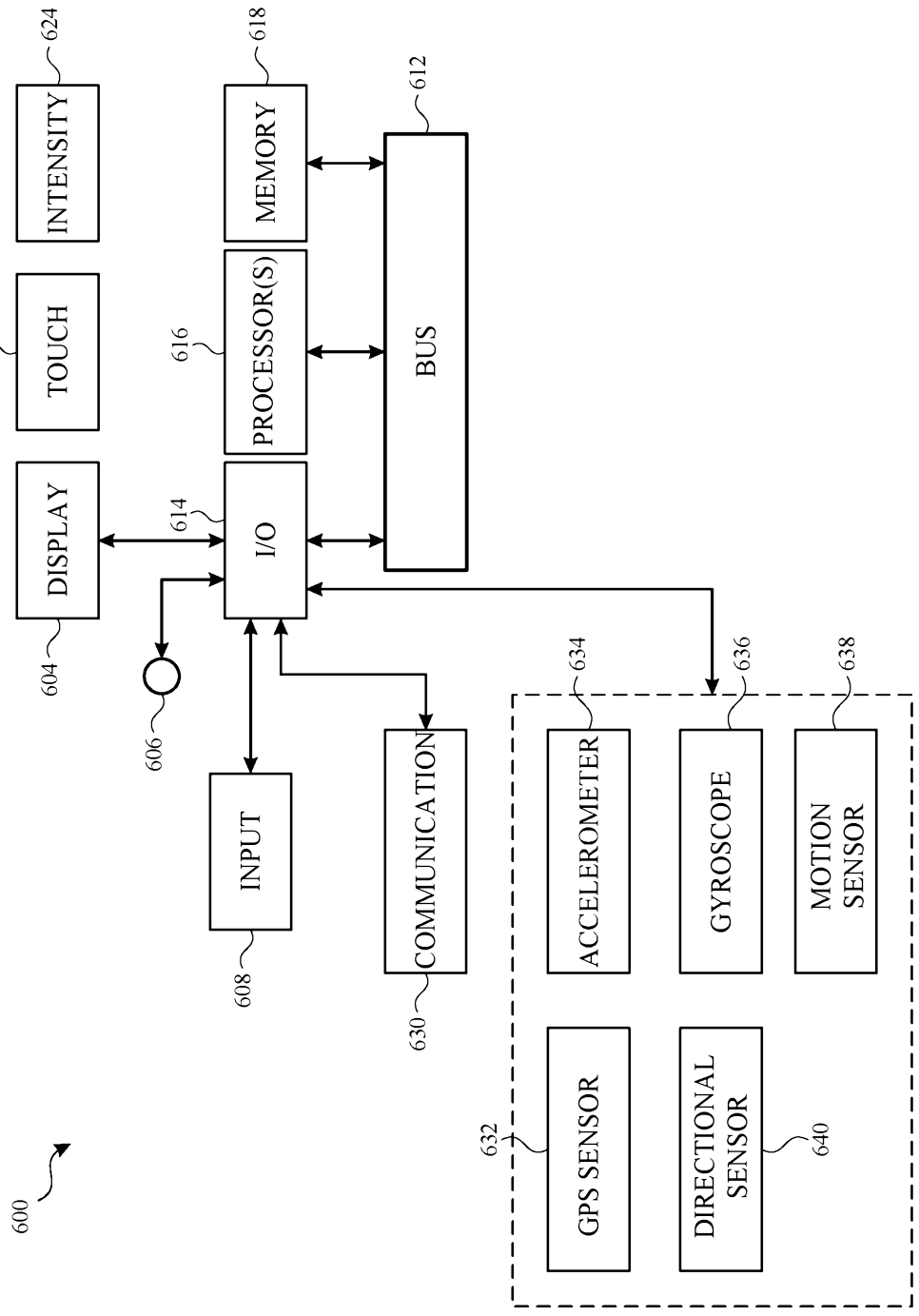

600
INTENSITY
624
TOUCH
622
DISPLAY
604
606
INPUT
608
I/O
614
PROCESSOR(S)
616
MEMORY
618
BUS
612
COMMUNICATION
630
ACCELEROMETER
634
GYROSCOPE
636
MOTION SENSOR
638
GPS SENSOR
632
DIRECTIONAL SENSOR
640

FIG. 6B

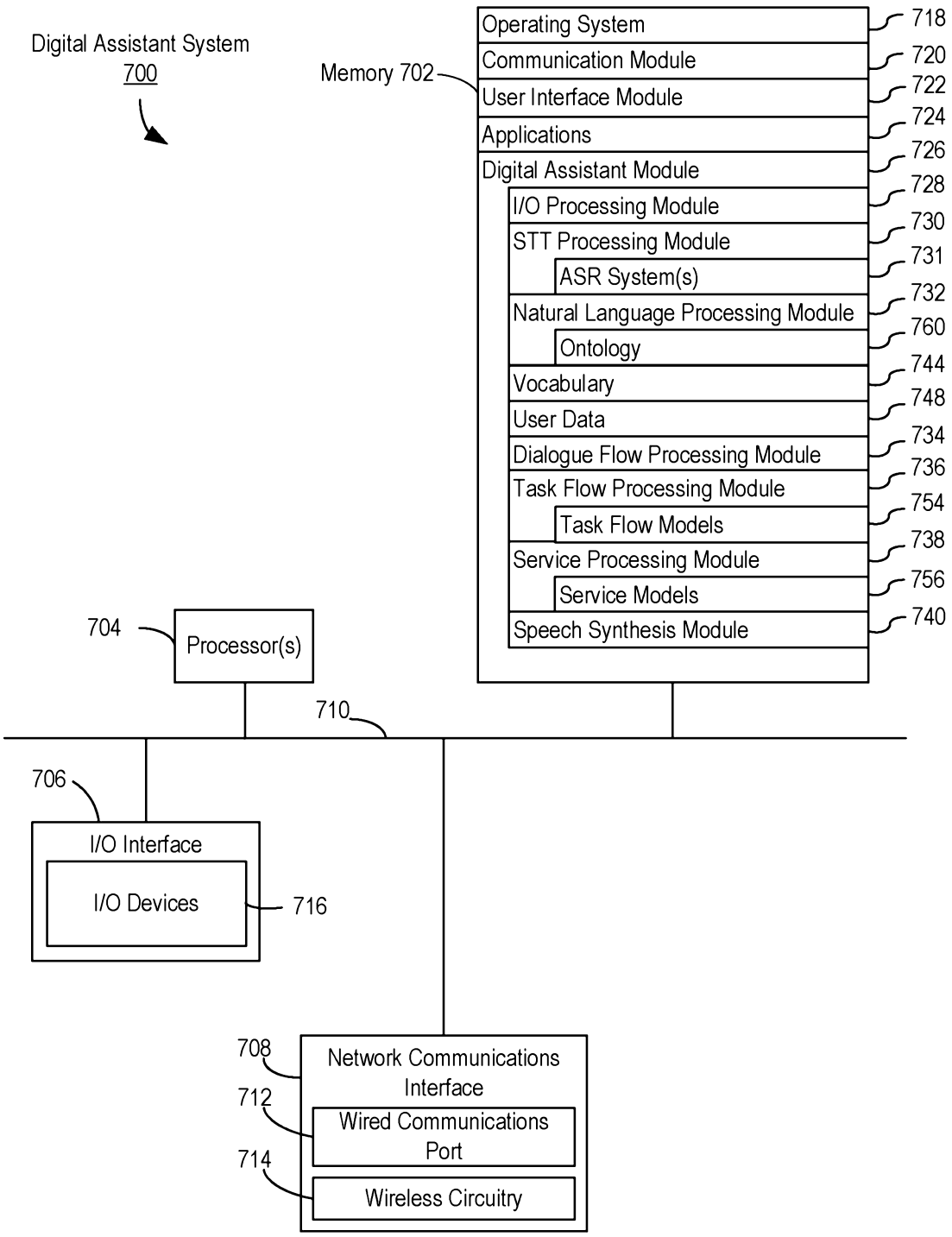

Digital Assistant System
700
Memory 702
Operating System
718
Communication Module
720
User Interface Module
722
Applications
724
Digital Assistant Module
726
I/O Processing Module
728
STT Processing Module
730
ASR System(s)
731
Natural Language Processing Module
732
Ontology
760
Vocabulary
744
User Data
748
Dialogue Flow Processing Module
734
Task Flow Processing Module
736
Task Flow Models
754
Service Processing Module
738
Service Models
756
Speech Synthesis Module
740
704
Processor(s)
710
706
I/O Interface
I/O Devices
716
708
Network Communications Interface
712
Wired Communications Port
714
Wireless Circuitry

FIG. 7A

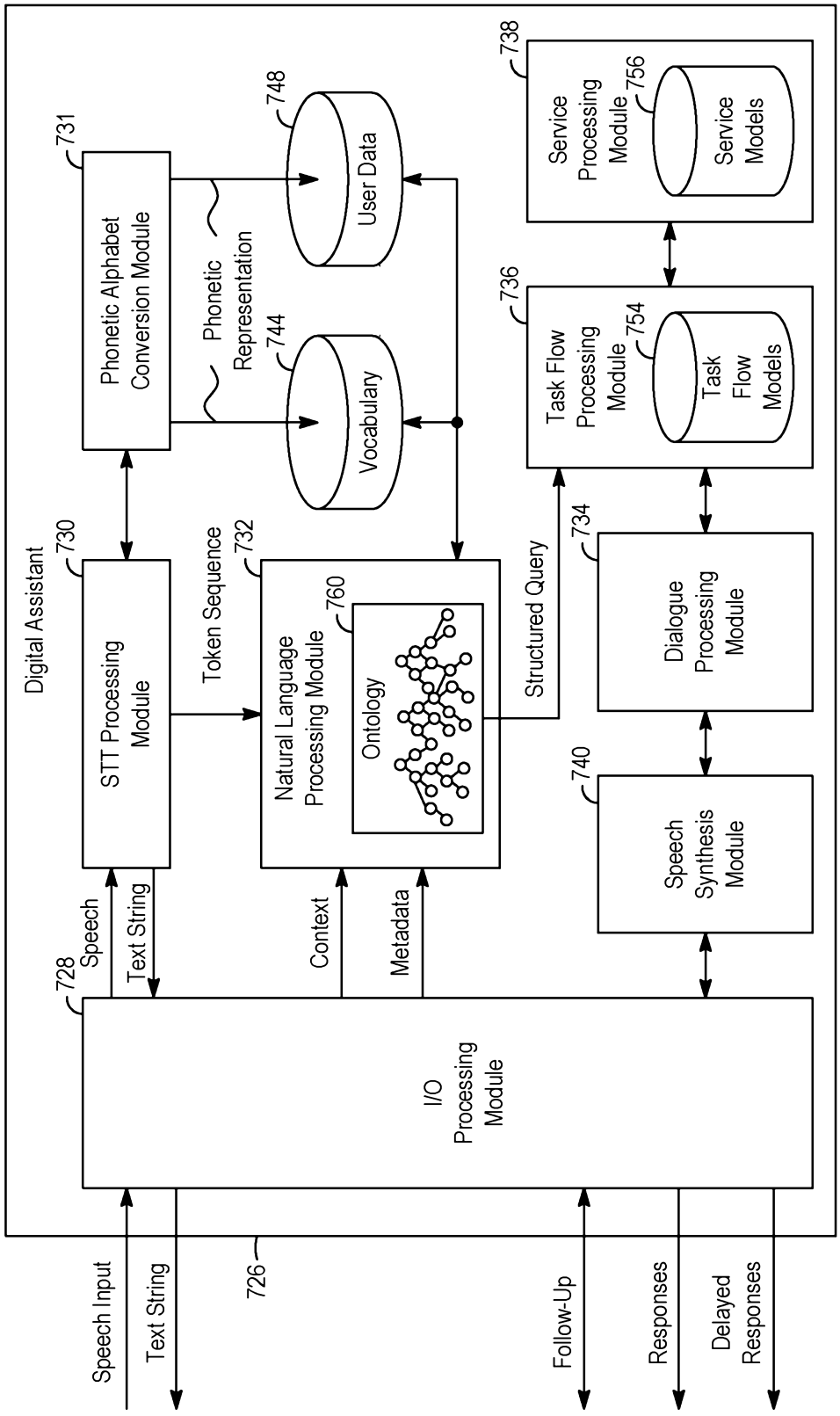

Digital Assistant
726
728
I/O Processing Module
Speech
Text String
730
STT Processing Module
731
Phonetic Alphabet Conversion Module
732
Natural Language Processing Module
760
Ontology
Token Sequence
Context
Metadata
734
Dialogue Processing Module
736
Task Flow Processing Module
754
Task Flow Models
738
Service Processing Module
756
Service Models
740
Speech Synthesis Module
744
Vocabulary
748
User Data
Phonetic Representation
Structured Query
Speech Input
Text String
Follow-Up
Responses
Delayed Responses

FIG. 7B

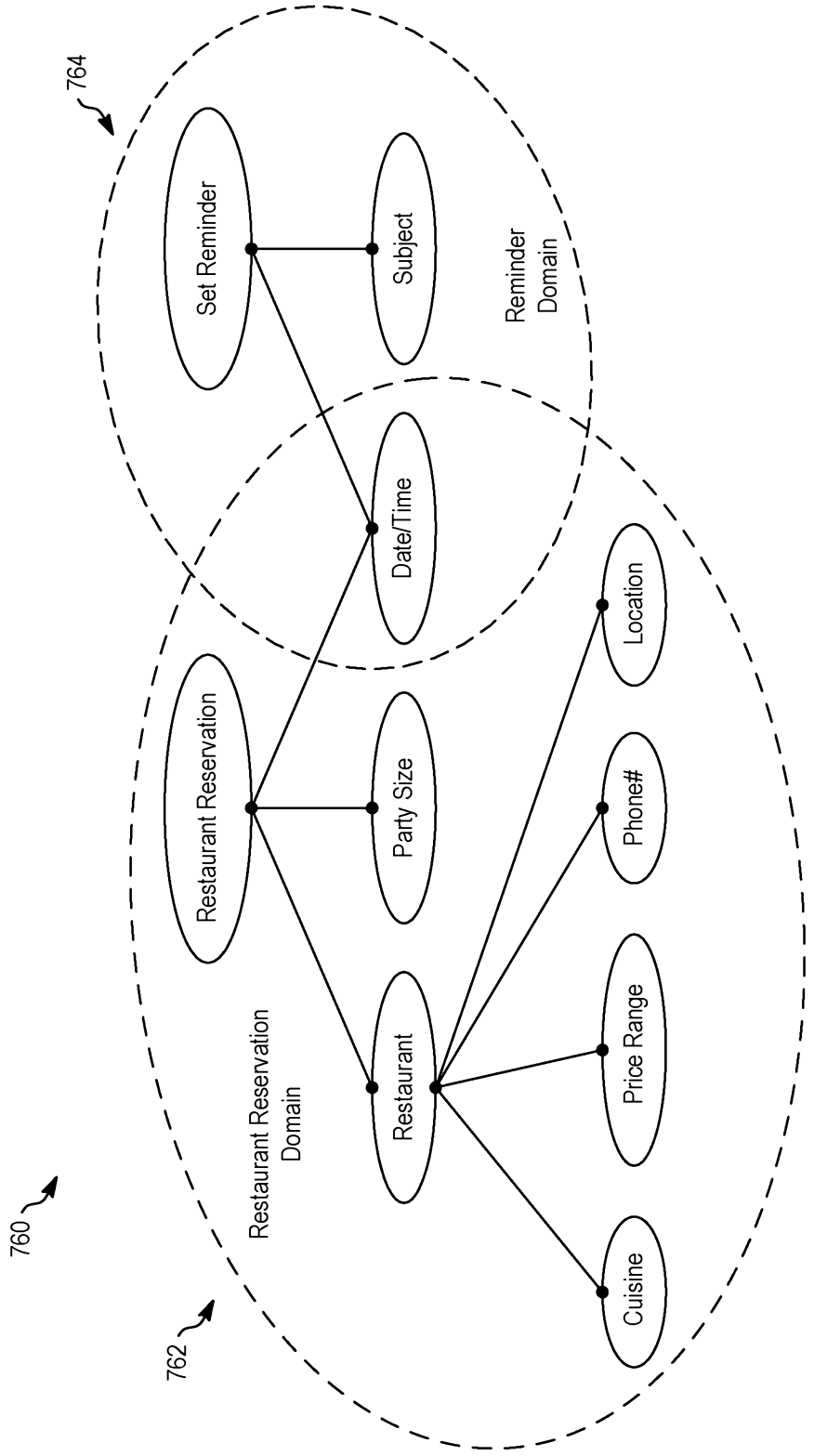

760
762
764
Restaurant Reservation Domain
Reminder Domain
Restaurant Reservation
Party Size
Restaurant
Cuisine
Price Range
Phone#
Location
Date/Time
Set Reminder
Subject

FIG. 7C

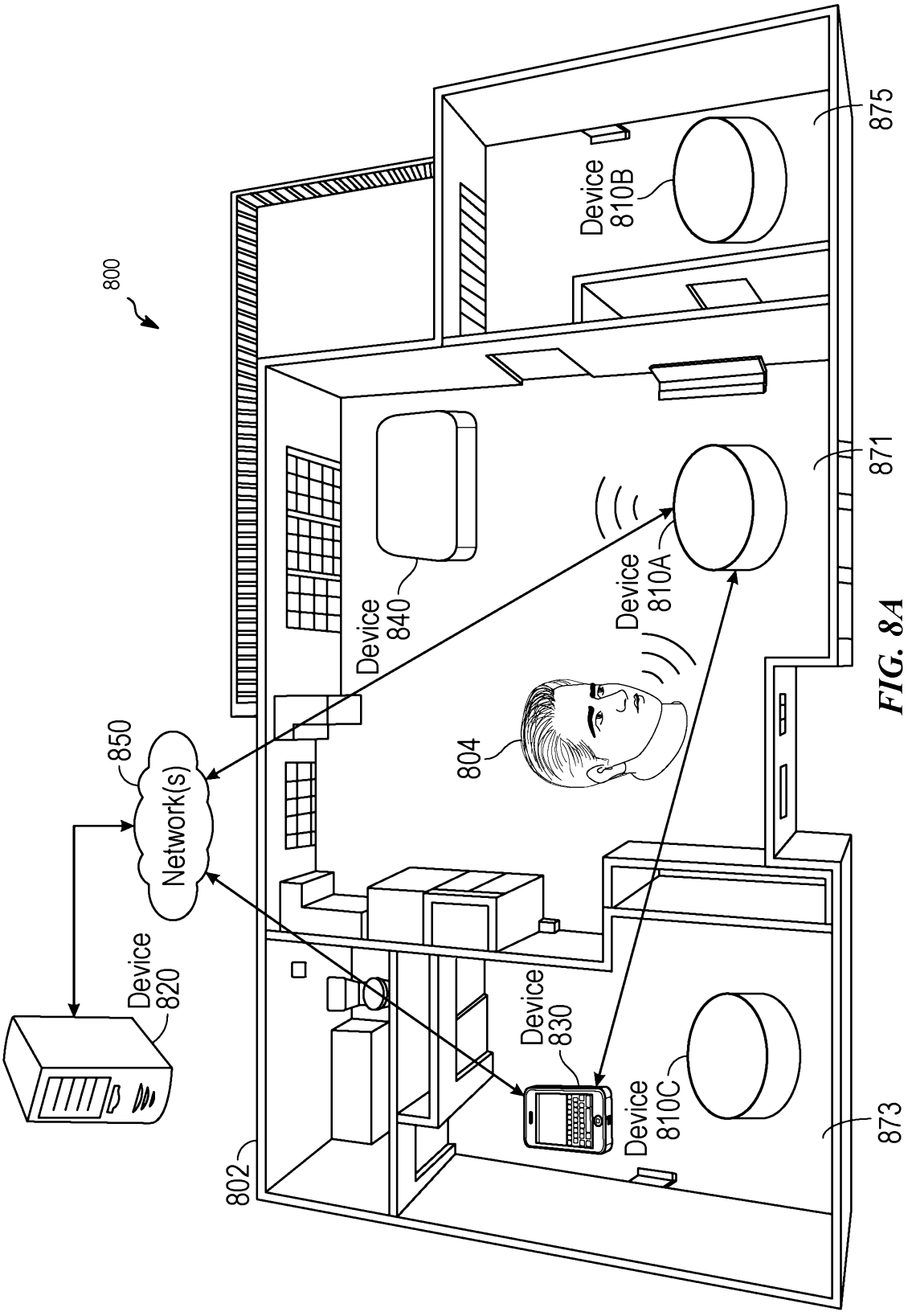

800
875
Device
810B
871
Device
840
Device
810A
804
850
Network(s)
Device
820
Device
830
Device
810C
873
802

FIG. 8A

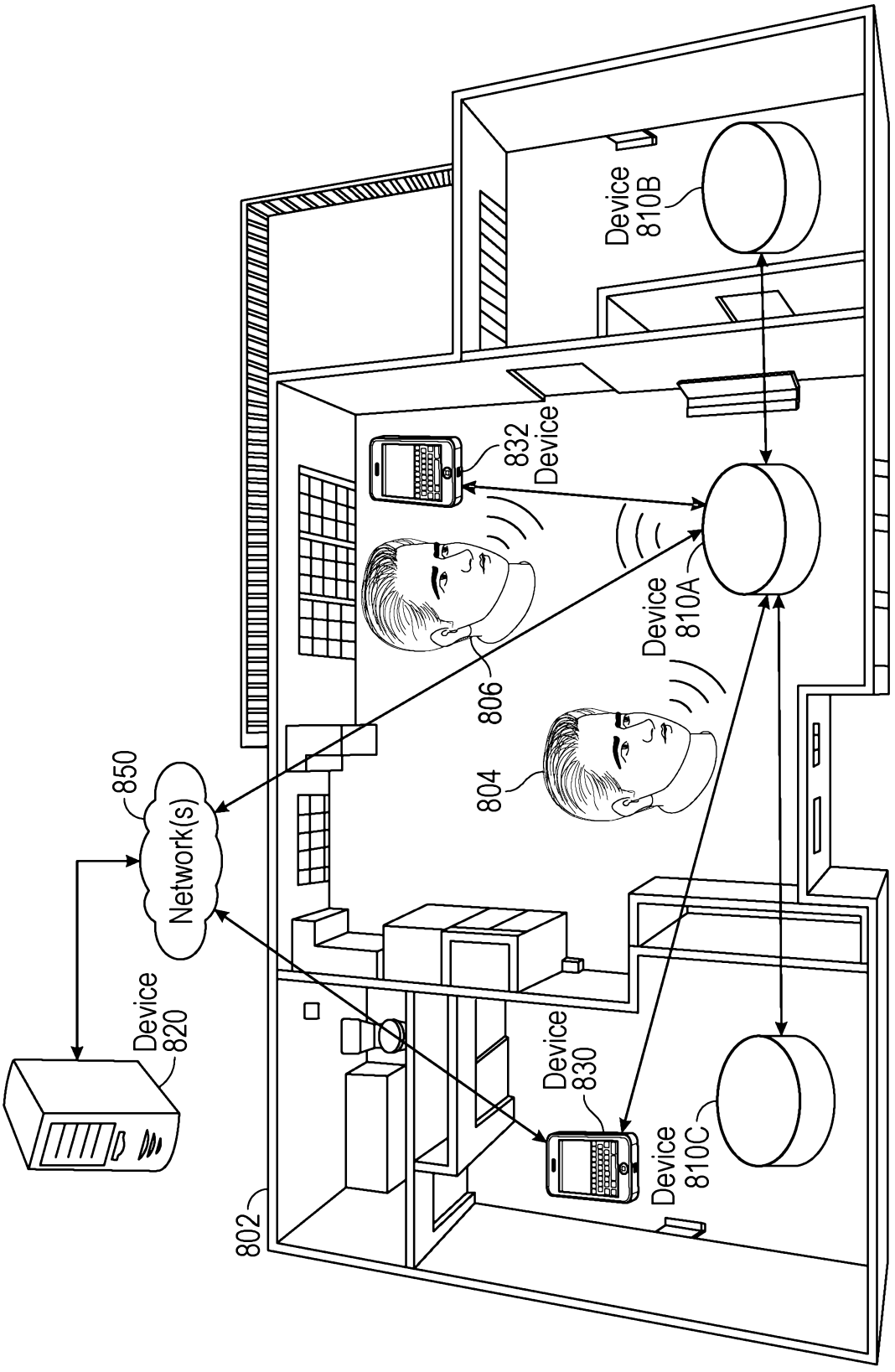

Device 810B
832 Device
Device 810A
806
804
850
Network(s)
Device 820
Device 830
Device 810C
802

FIG. 8B

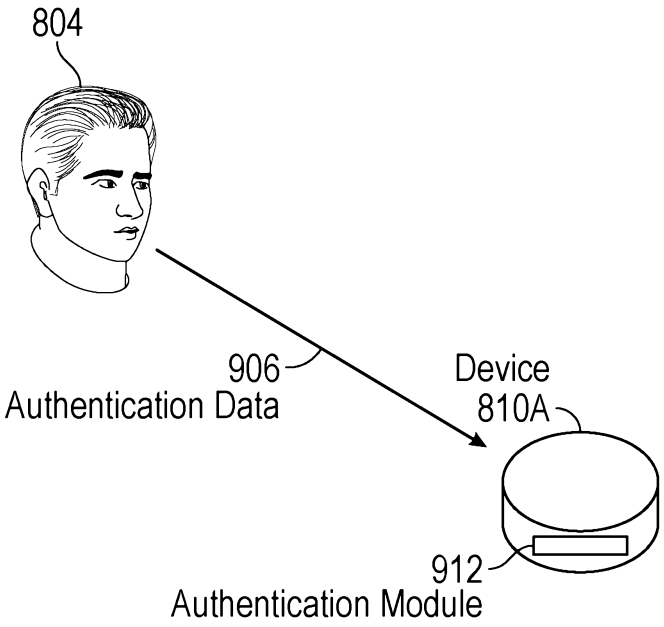

804
906
Authentication Data
Device
810A
912
Authentication Module

FIG. 9A

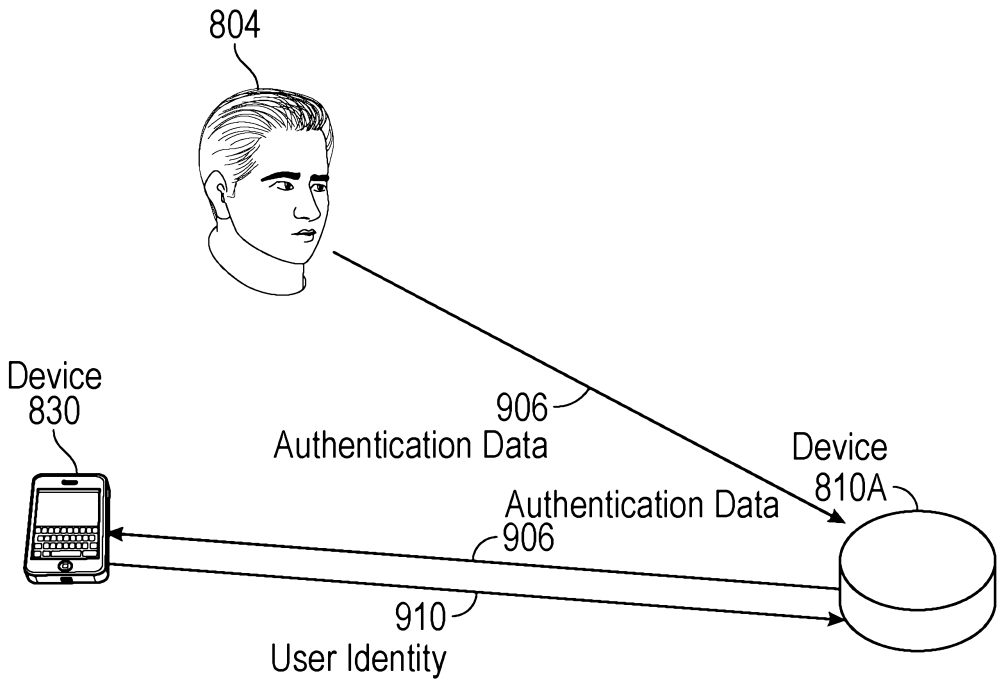

804
Device
830
906
Authentication Data
Device
810A
Authentication Data
906
910
User Identity

FIG. 9B

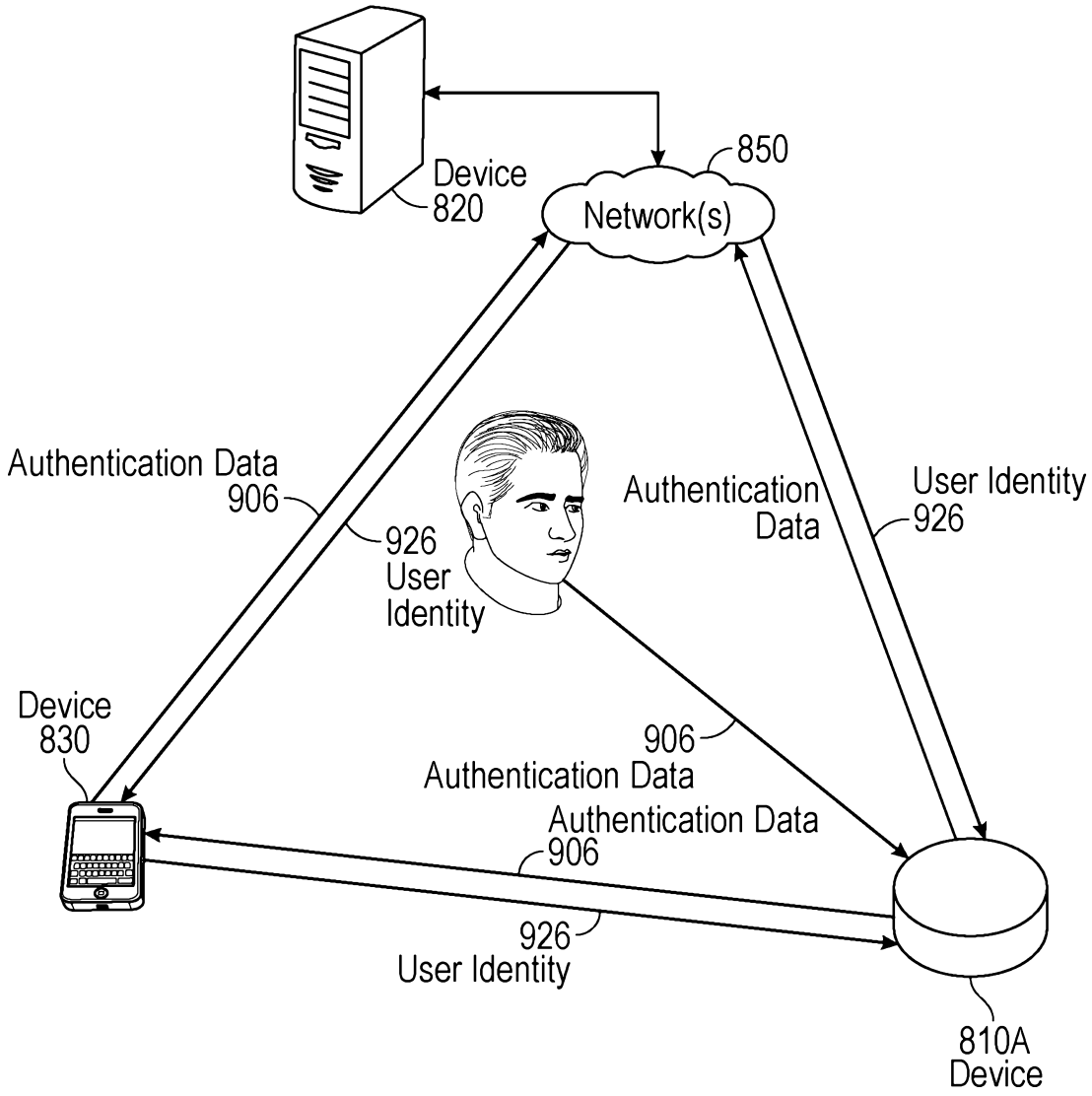

Device
820
850
Network(s)
Authentication Data
906
926
User
Identity
Authentication
Data
User Identity
926
Device
830
906
Authentication Data
Authentication Data
906
926
User Identity
810A
Device

FIG. 9C

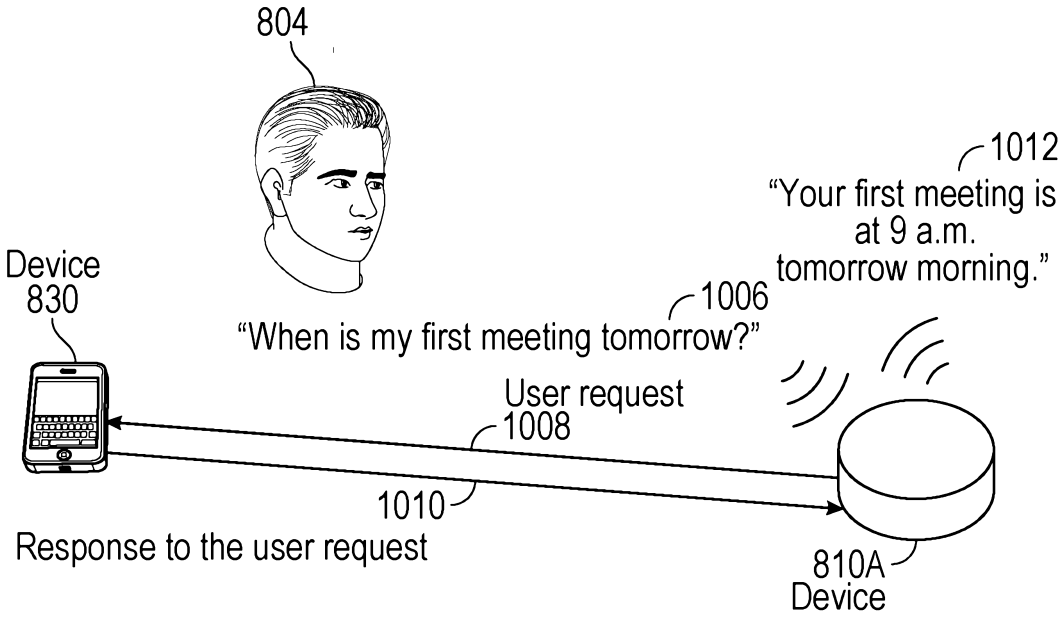

804
1012
"Your first meeting is at 9 a.m. tomorrow morning."
Device 830
1006
"When is my first meeting tomorrow?"
User request
1008
1010
Response to the user request
810A
Device

FIG. 10A

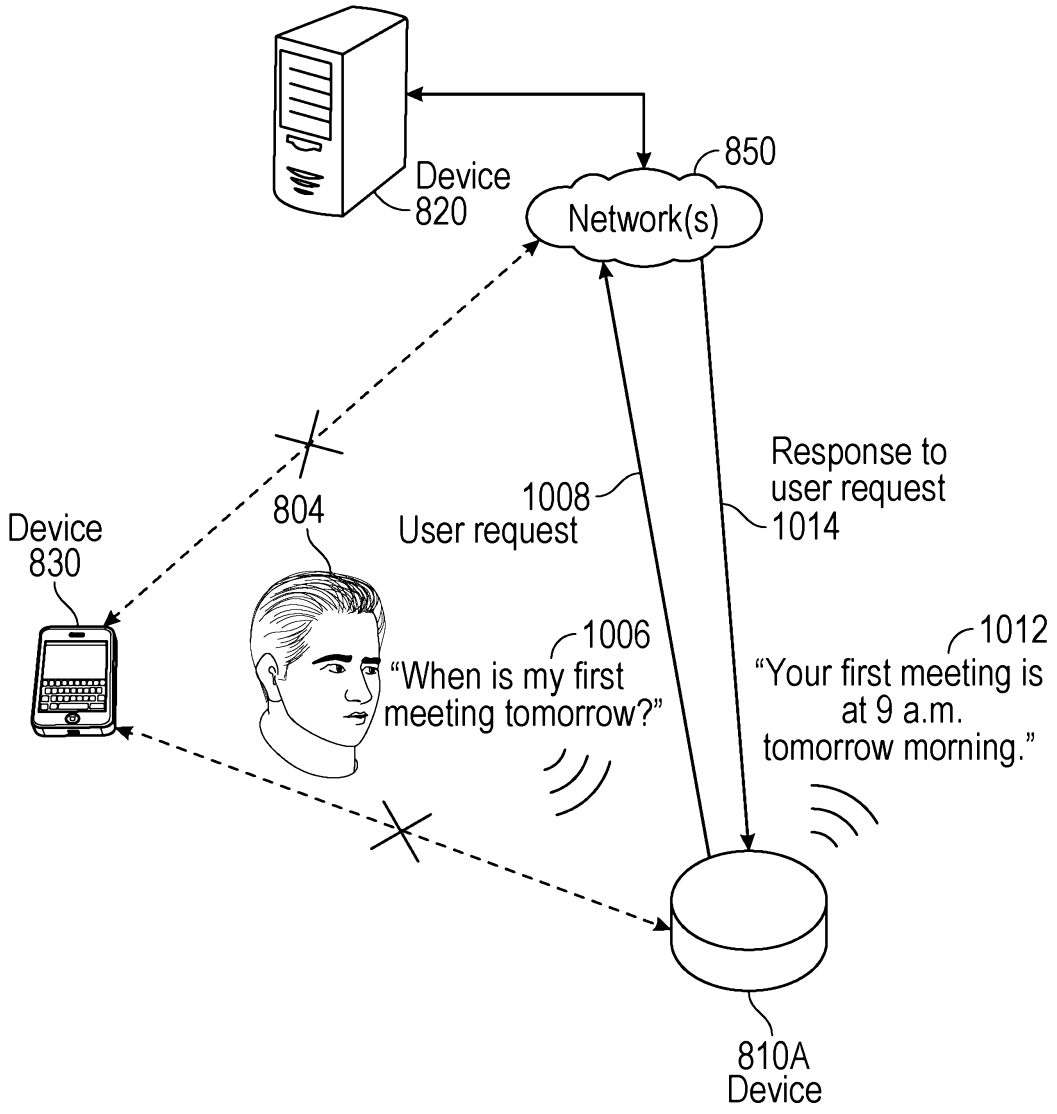

Device
820
850
Network(s)
804
1008
User request
Response to
user request
1014
Device
830
1006
"When is my first
meeting tomorrow?"
1012
"Your first meeting is
at 9 a.m.
tomorrow morning."
810A
Device

FIG. 10B

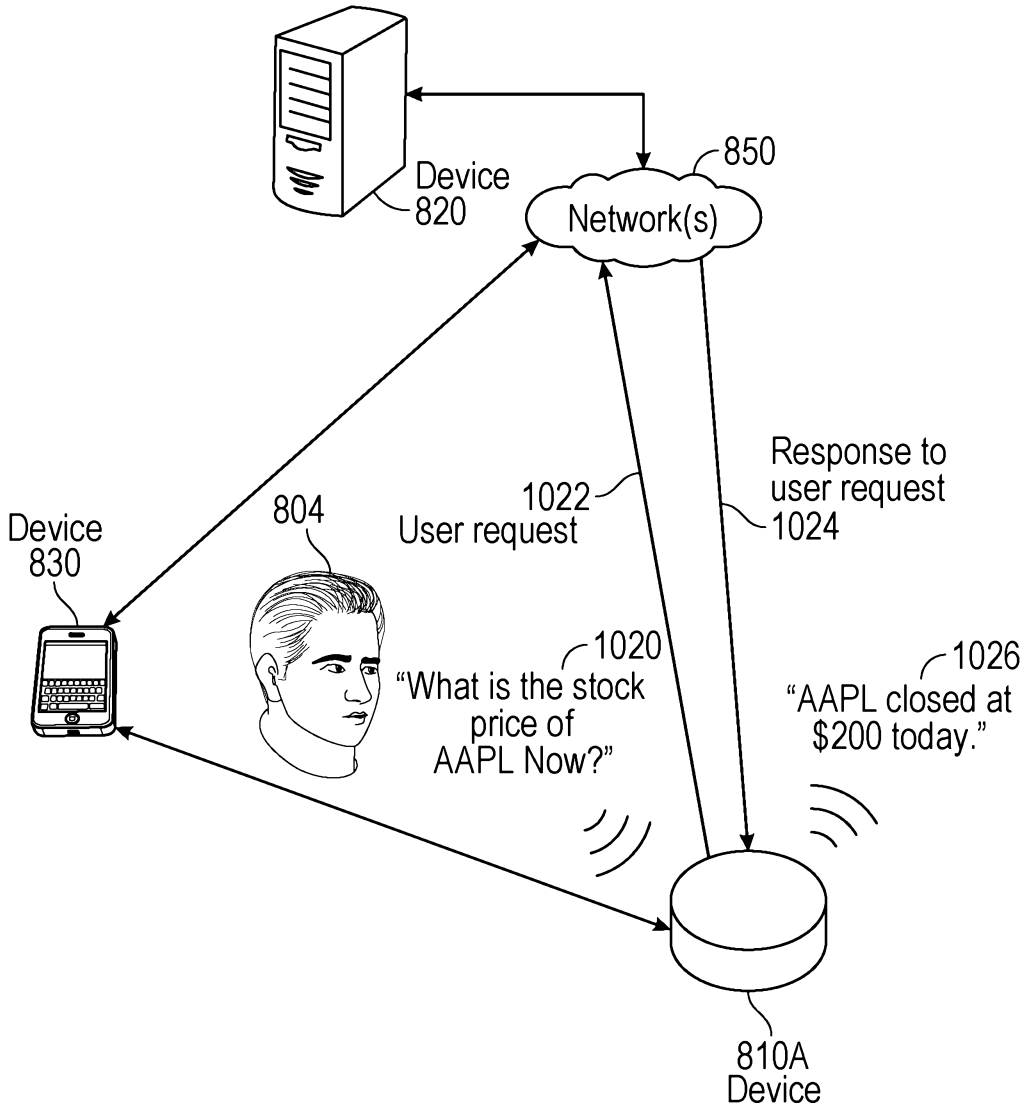

Device
820
850
Network(s)
Response to
user request
1024
1022
User request
804
Device
830
1020
"What is the stock
price of
AAPL Now?"
1026
"AAPL closed at
$200 today."
810A
Device

FIG. 10C

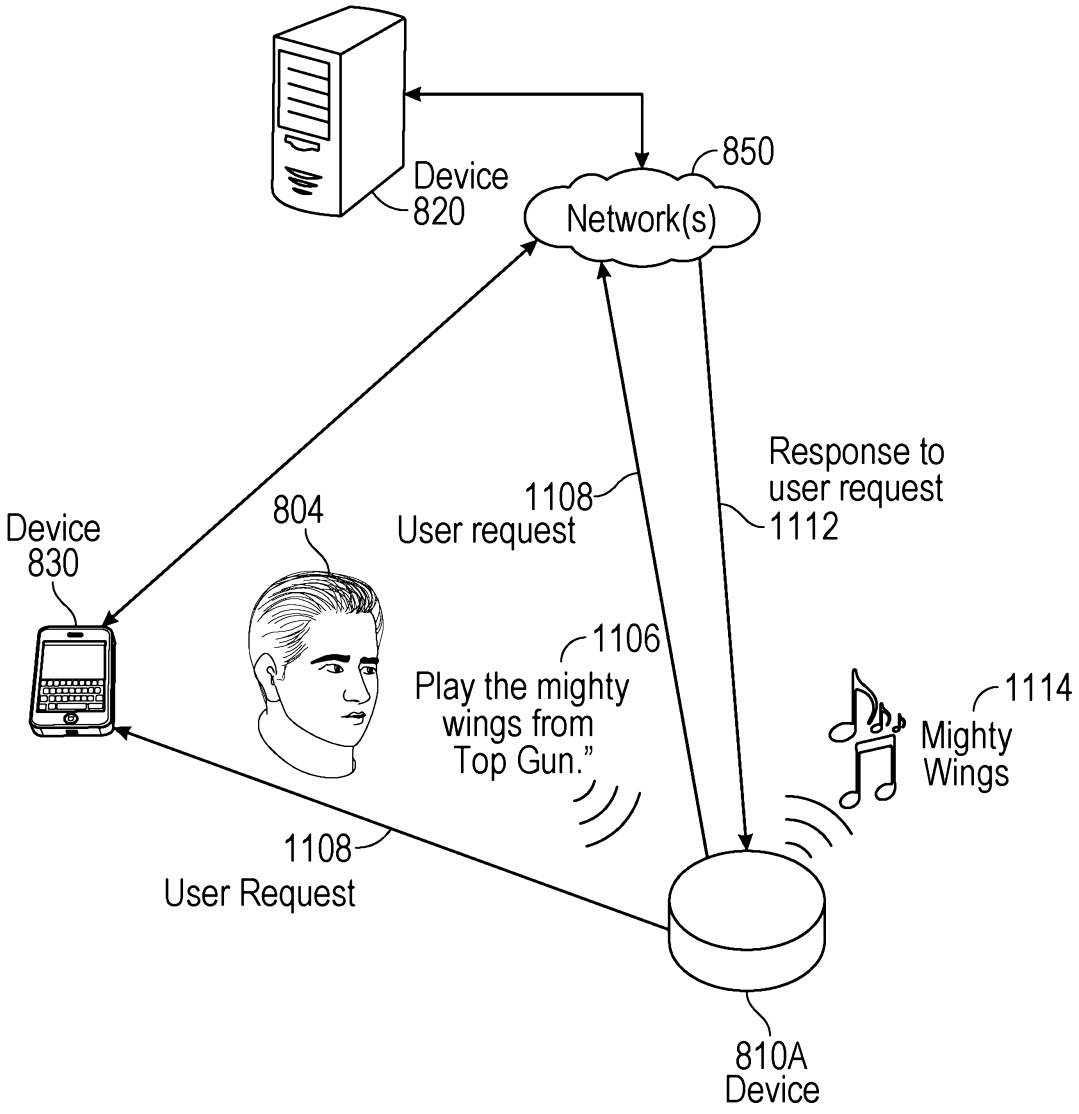

Device
820
850
Network(s)
Response to
user request
1112
1108
User request
804
Device
830
1106
Play the mighty
wings from
Top Gun."
1114
Mighty
Wings
1108
User Request
810A
Device

FIG. 11A

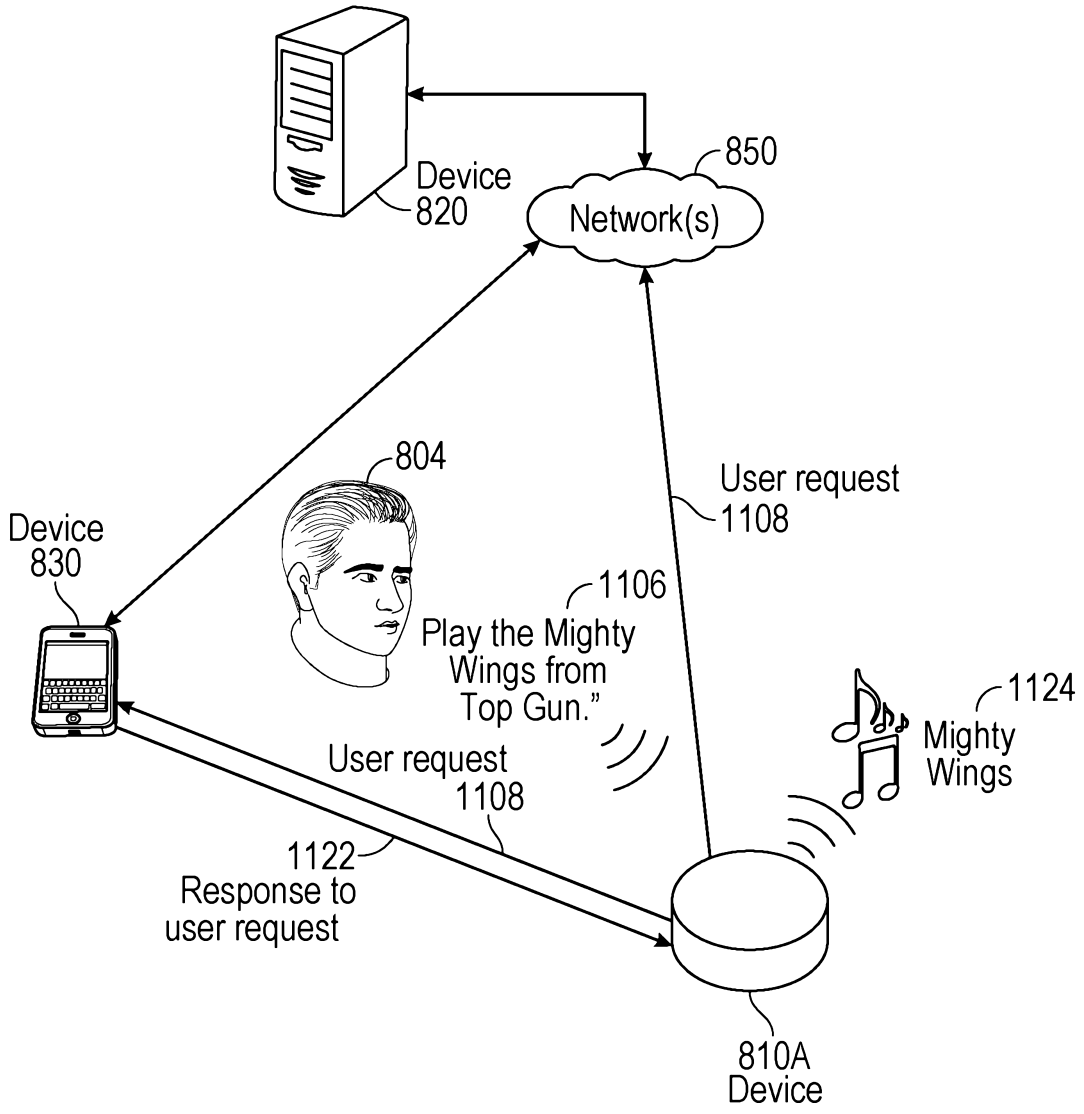

Device
820
850
Network(s)
804
User request
1108
Device
830
1106
Play the Mighty
Wings from
Top Gun."
1124
Mighty
Wings
User request
1108
1122
Response to
user request
810A
Device

FIG. 11B

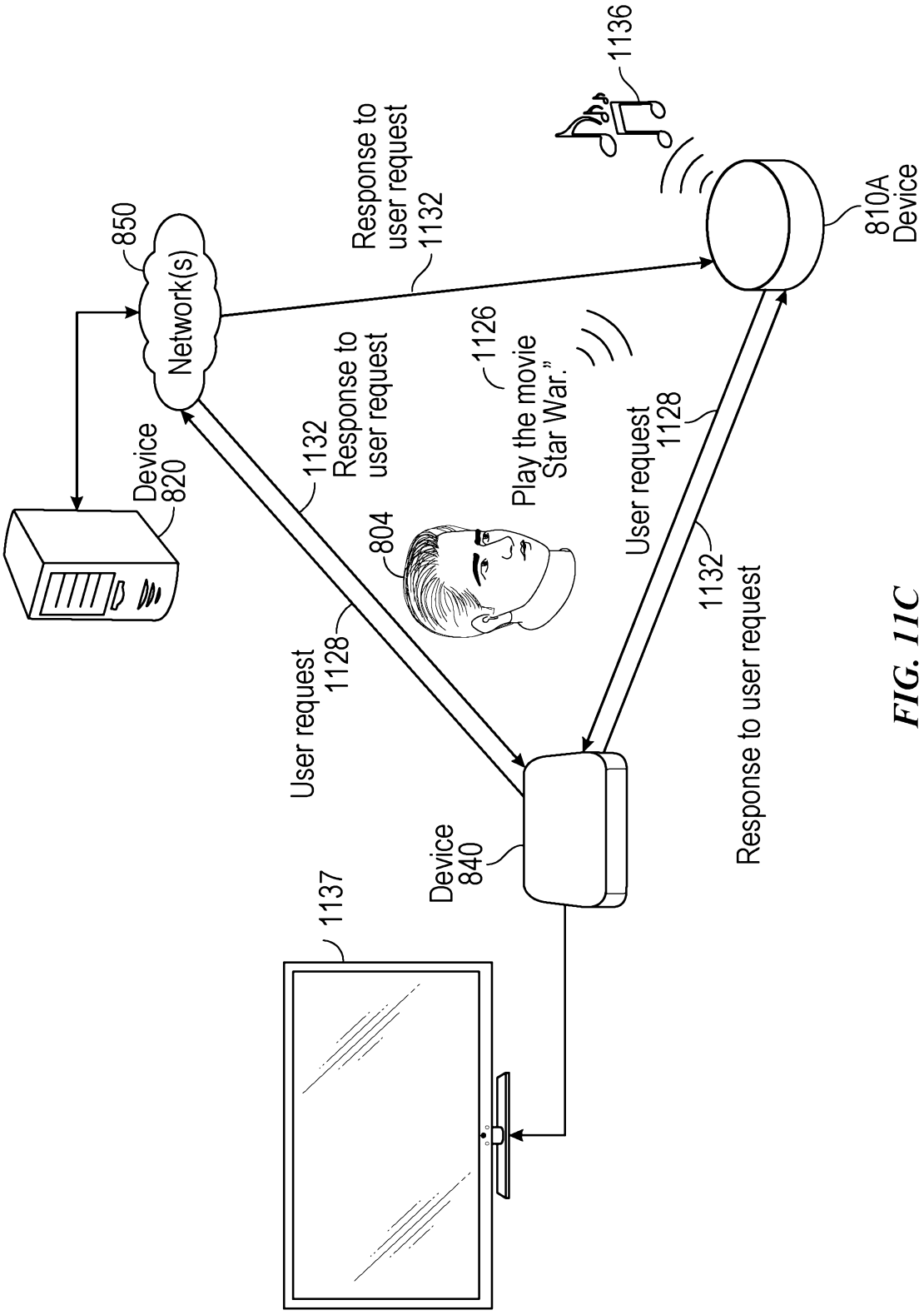

1136
Response to user request 1132
850
Network(s)
810A Device
1126
Play the movie Star War."
Response to user request 1132
User request 1128
Device 820
804
User request 1128
1132 Response to user request
Device 840
1137

FIG. 11C

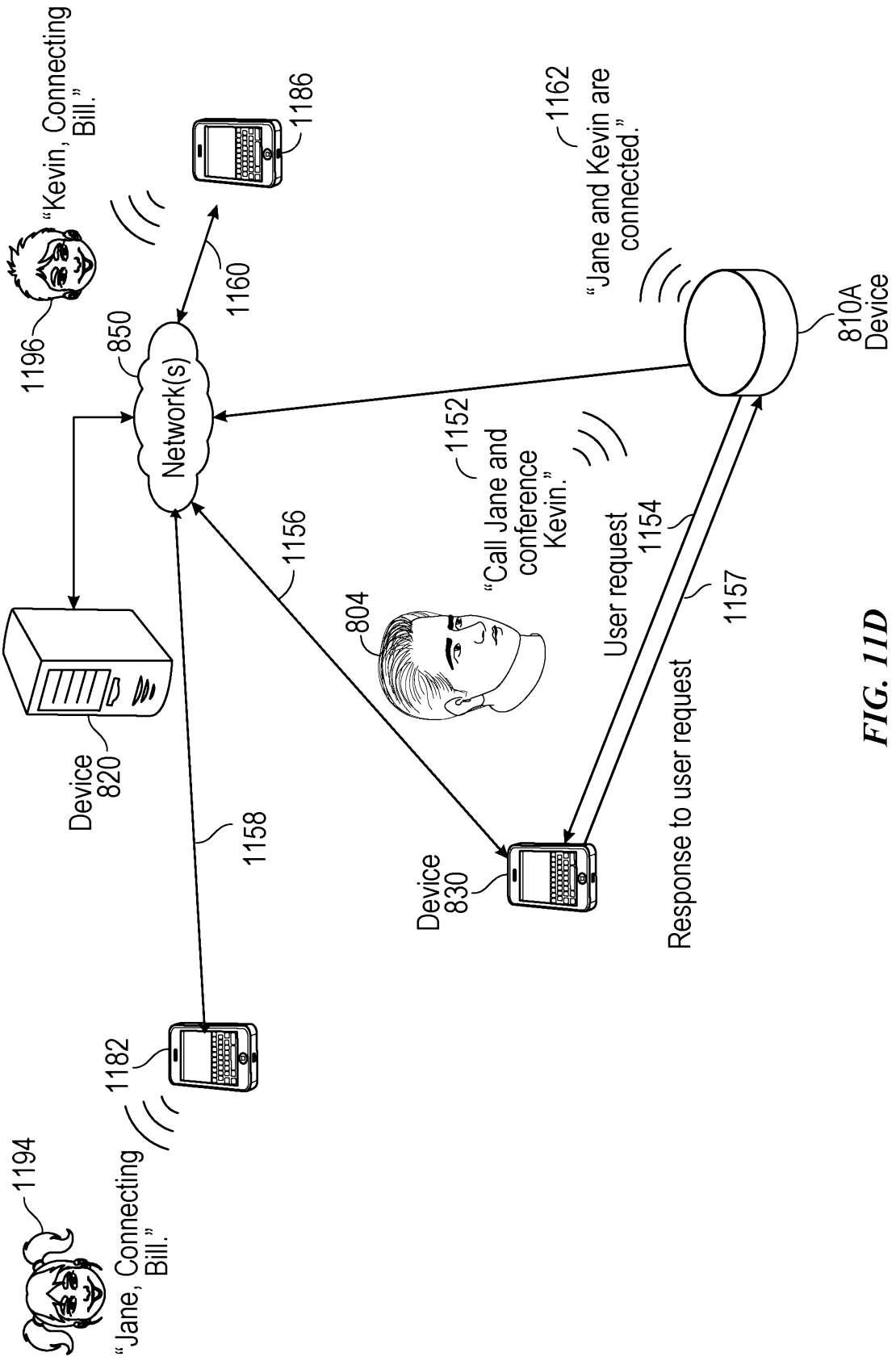

"Kevin, Connecting Bill."
1196
1186
1160
850
Network(s)
1162
"Jane and Kevin are connected."
810A Device
1152
"Call Jane and conference Kevin."
User request 1154
1157
Response to user request
1156
804
Device 820
1158
Device 830
1182
1194
"Jane, Connecting Bill."

FIG. 11D

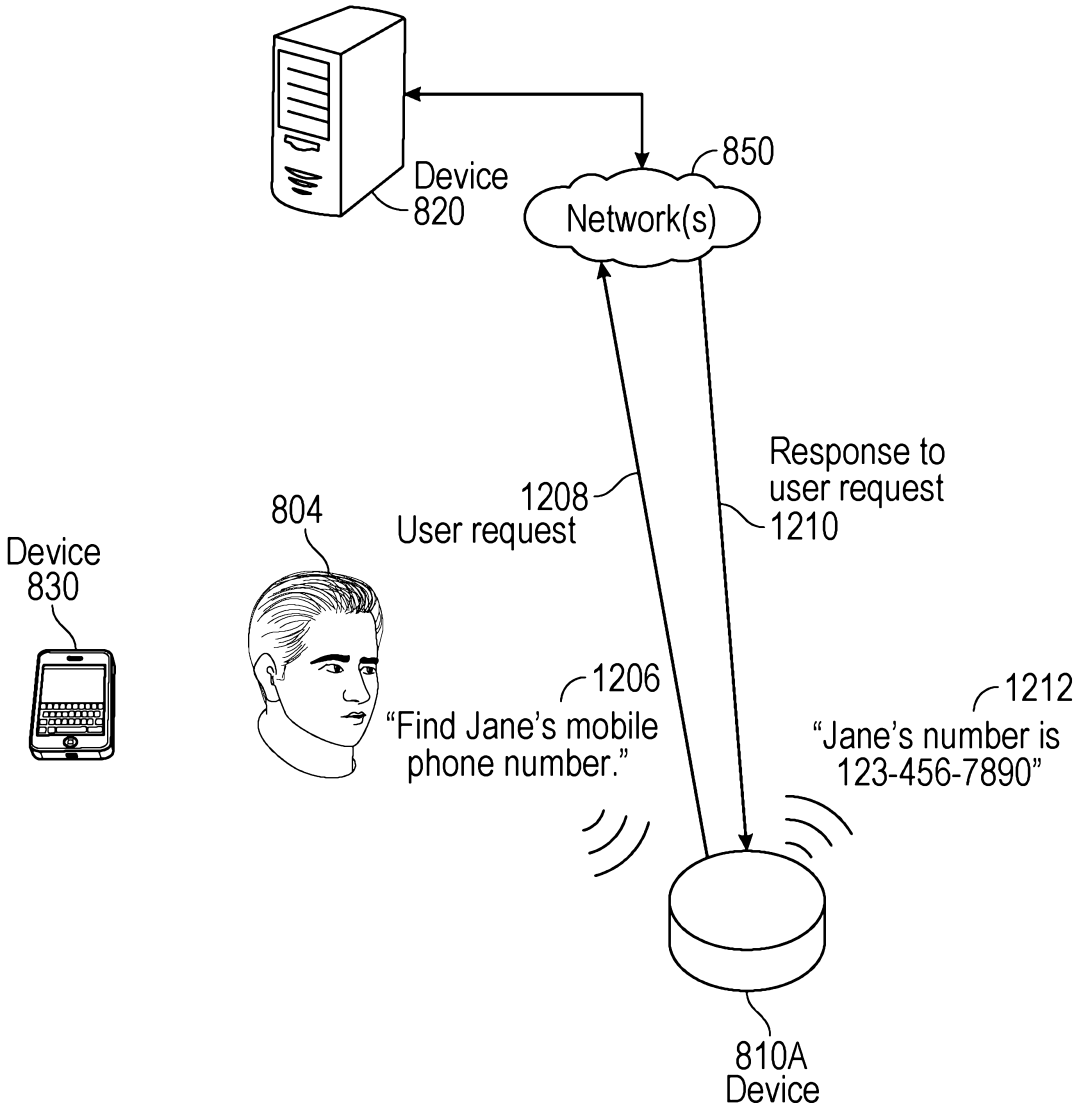

Device
820
850
Network(s)
Response to
user request
1210
1208
User request
804
Device
830
1206
"Find Jane's mobile
phone number."
1212
"Jane's number is
123-456-7890"
810A
Device

FIG. 12A

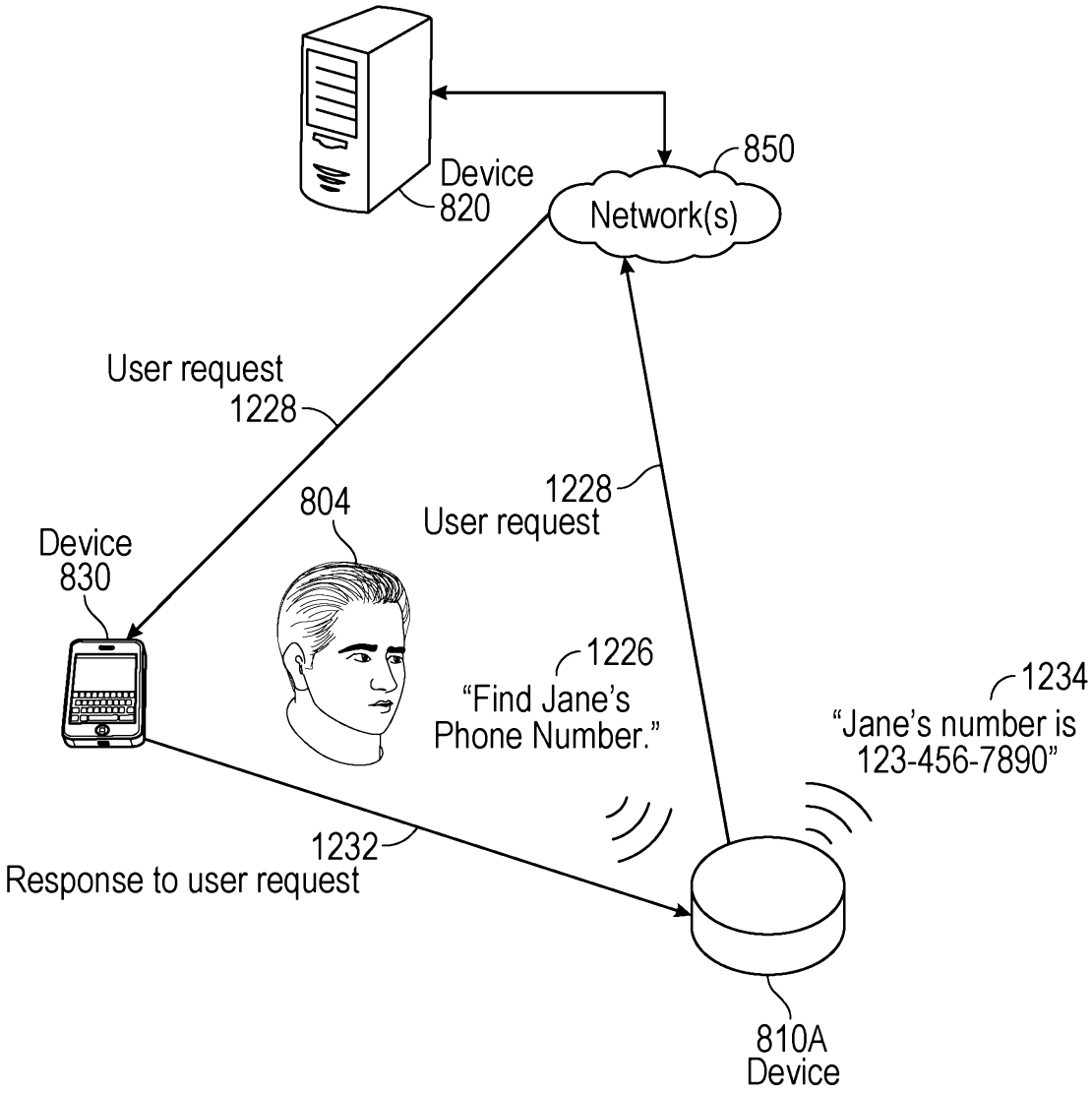

Device
820
850
Network(s)
User request
1228
804
1228
User request
Device
830
1226
"Find Jane's
Phone Number."
1234
"Jane's number is
123-456-7890"
1232
Response to user request
810A
Device

FIG. 12B

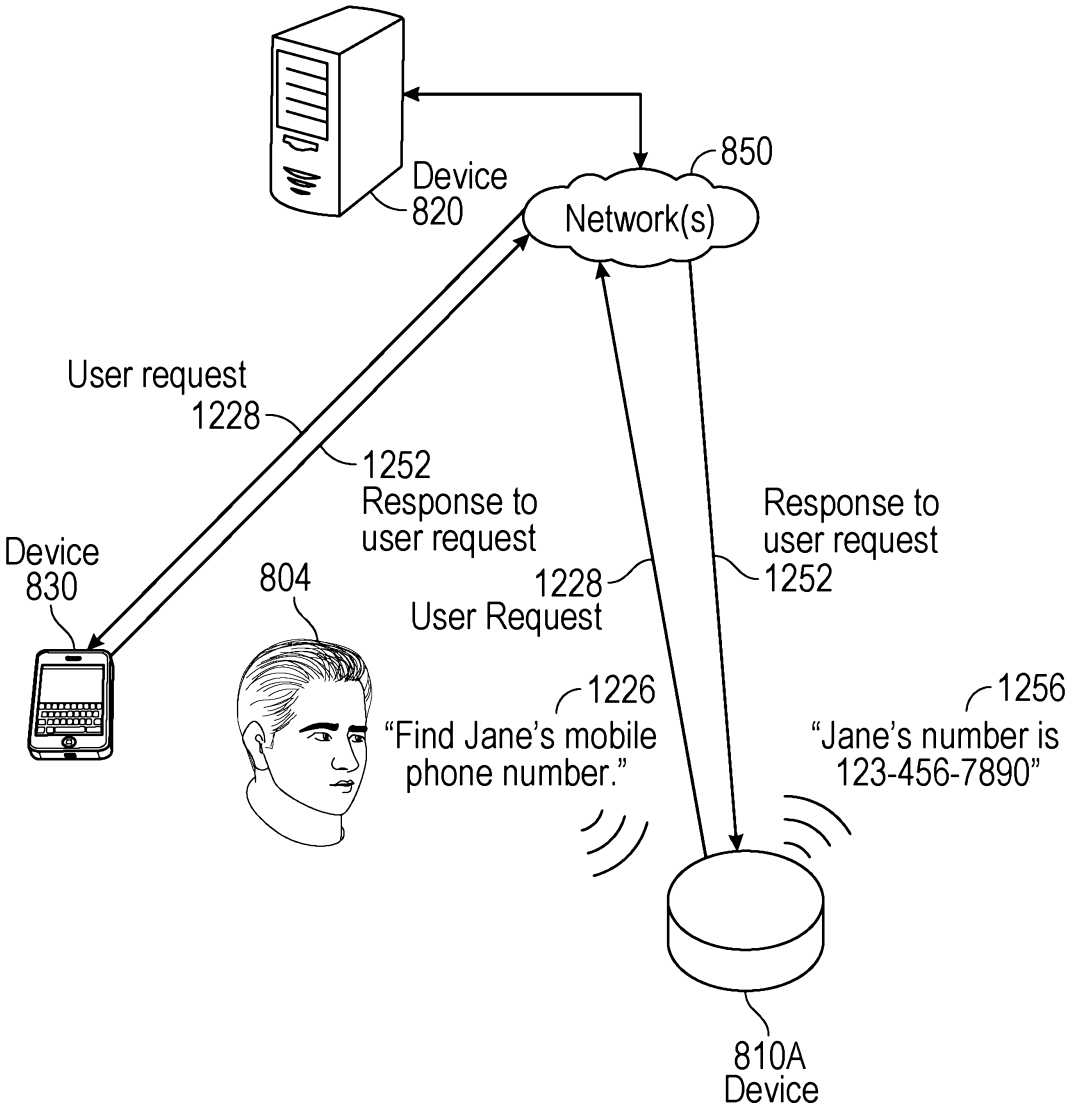

Device
820
850
Network(s)
User request
1228
1252
Response to
user request
Device
830
804
1228
User Request
Response to
user request
1252
1226
"Find Jane's mobile
phone number."
1256
"Jane's number is
123-456-7890"
810A
Device

FIG. 12C

DK 2017 70434 A1

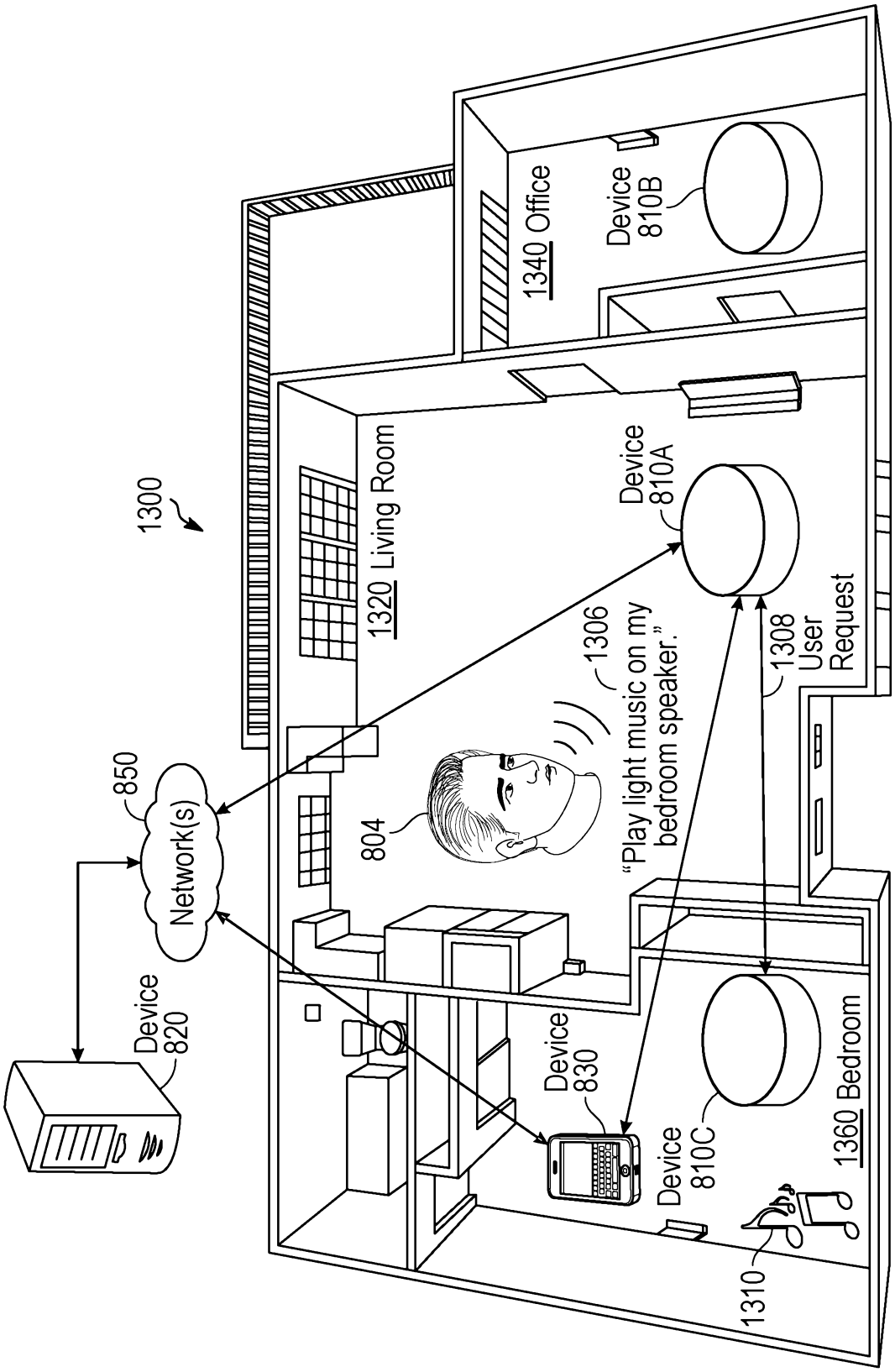


FIG. 13A

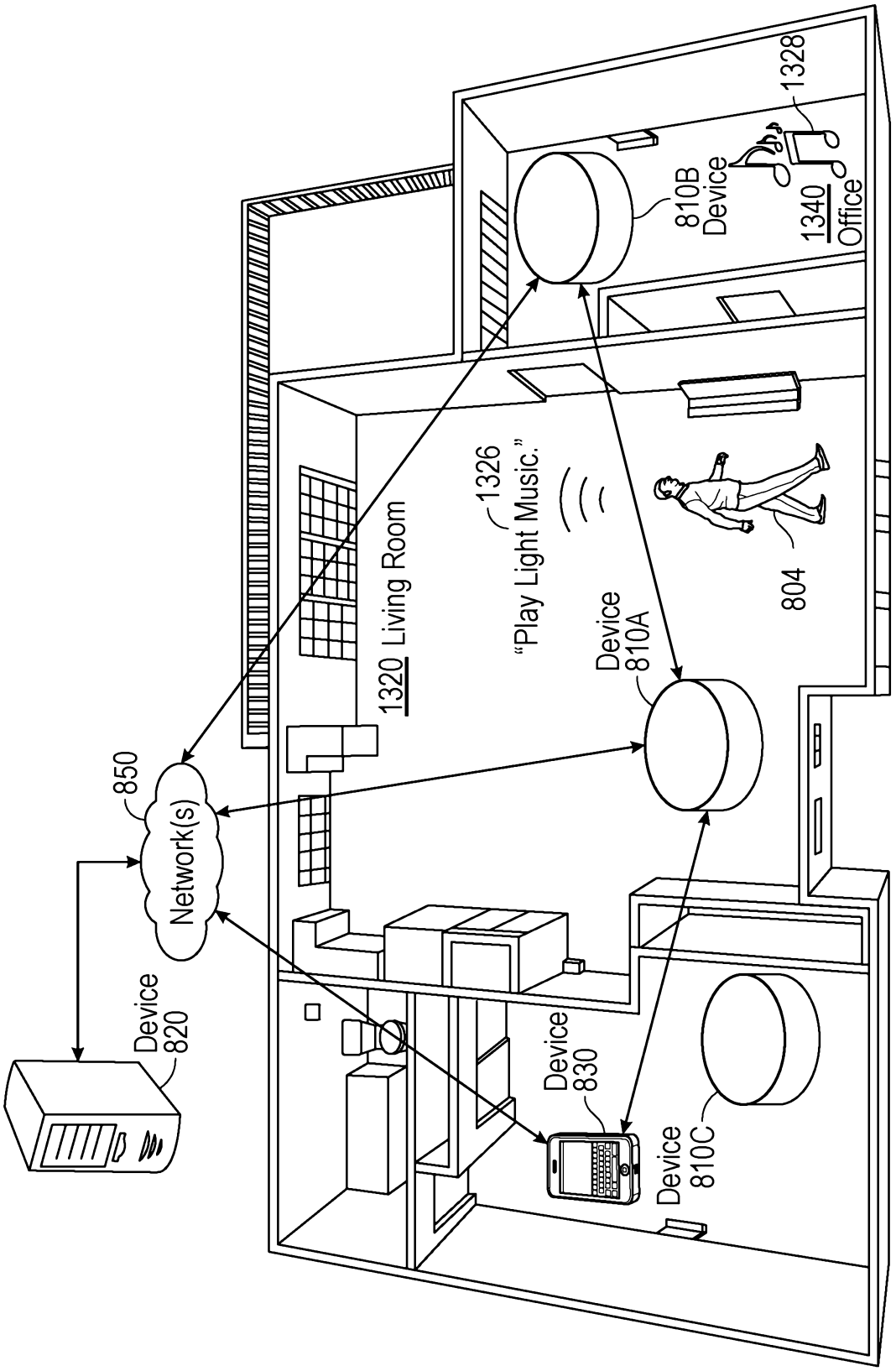

1328
810B Device
1340 Office
1326
"Play Light Music."
1320 Living Room
804
Device 810A
850
Network(s)
Device 820
Device 830
Device 810C

FIG. 13B

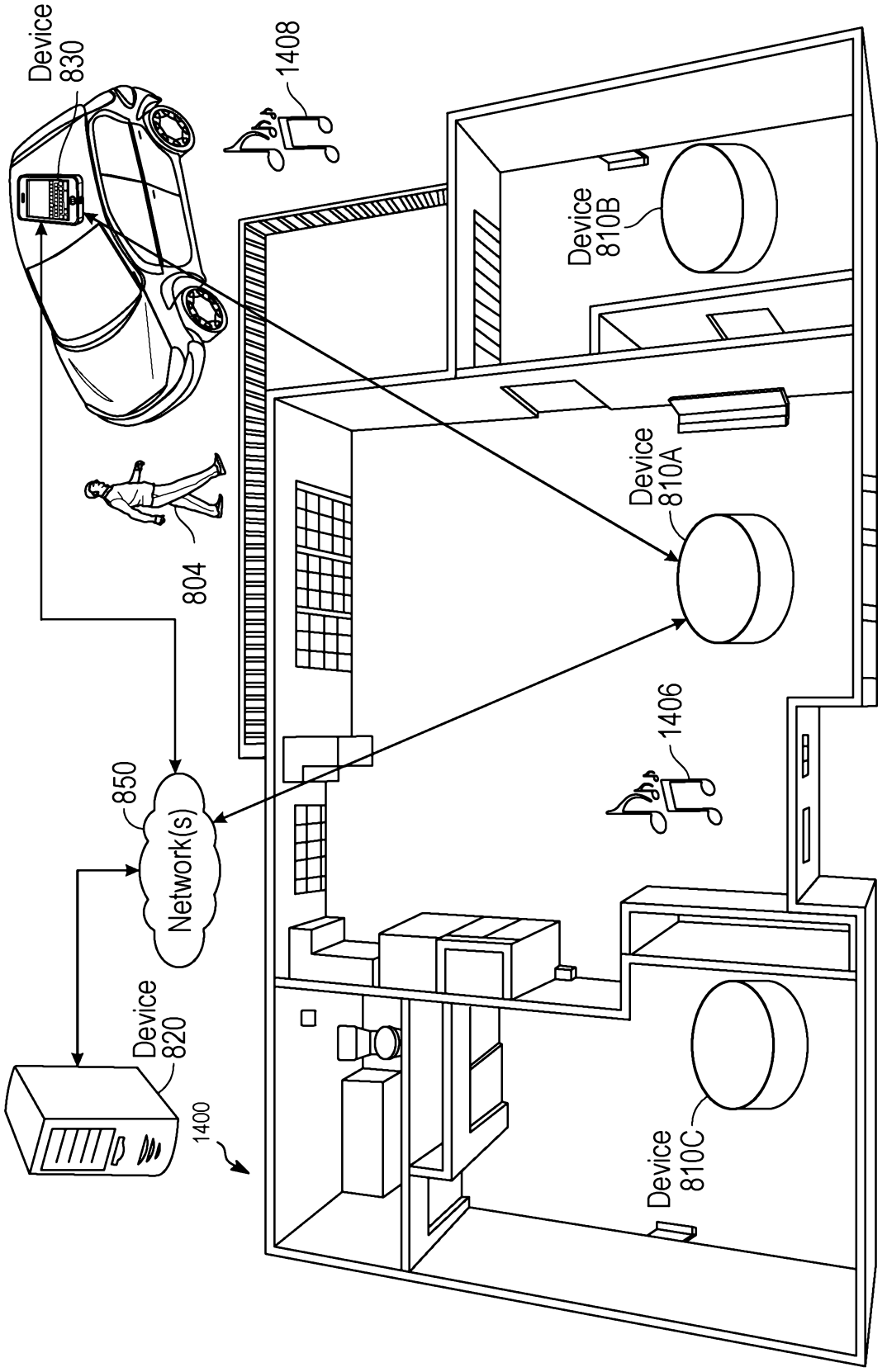

Device
830
1408
Device
810B
Device
810A
804
1406
850
Network(s)
Device
820
1400
Device
810C

FIG. 14

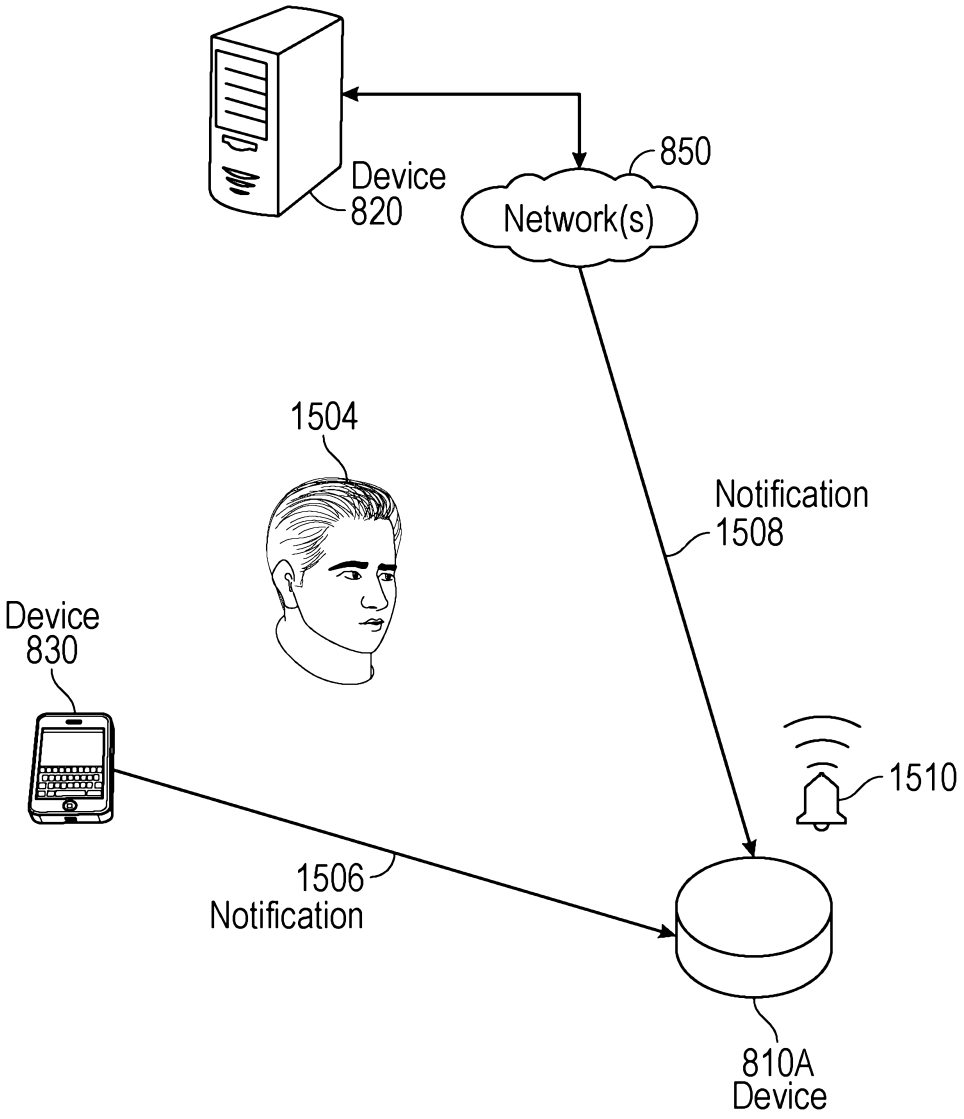

Device
820
850
Network(s)
1504
Notification
1508
Device
830
1510
1506
Notification
810A
Device

FIG. 15A

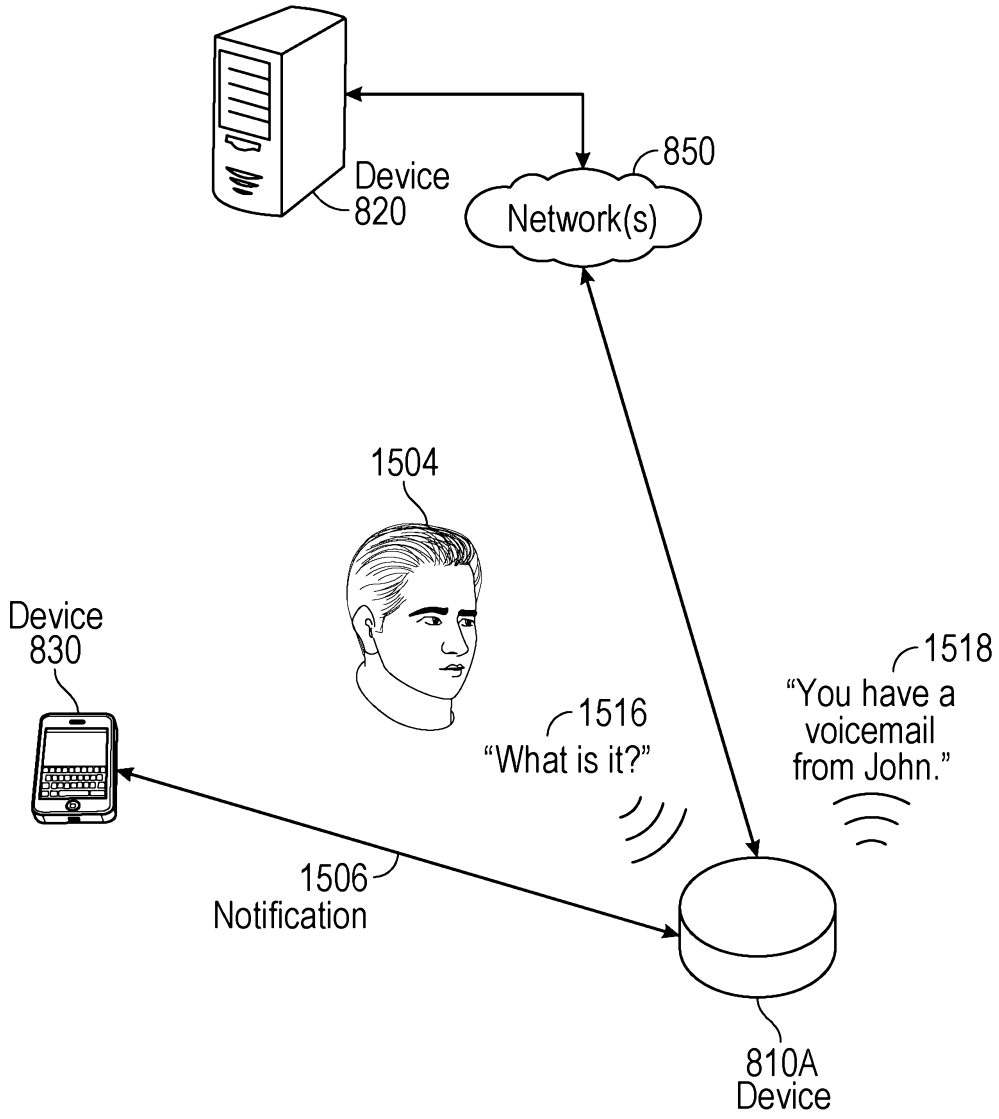

Device
820
850
Network(s)
1504
Device
830
1516
"What is it?"
1518
"You have a
voicemail
from John."
1506
Notification
810A
Device

FIG. 15B

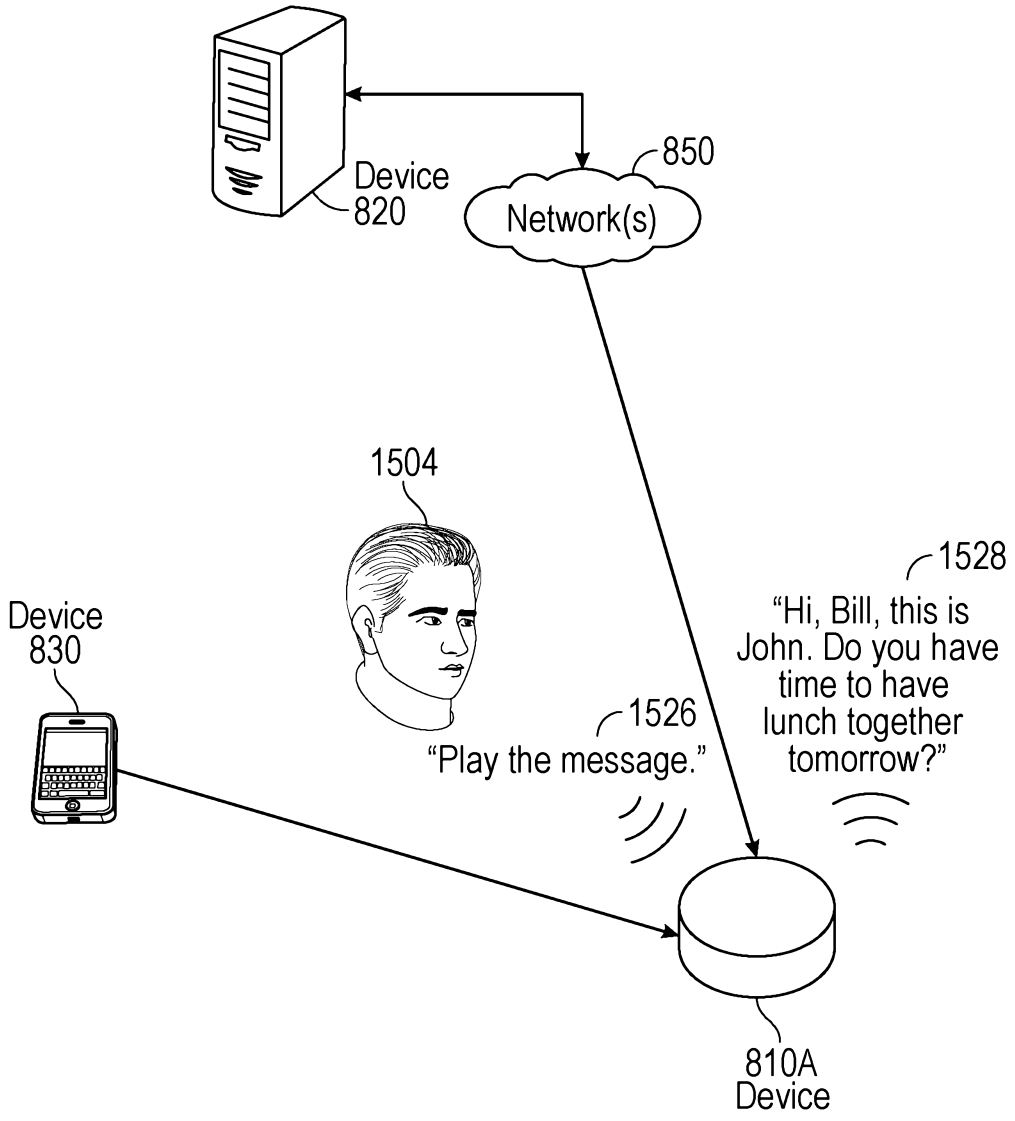

Device
820
850
Network(s)
1504
1528
"Hi, Bill, this is John. Do you have time to have lunch together tomorrow?"
Device
830
1526
"Play the message."
810A
Device

FIG. 15C

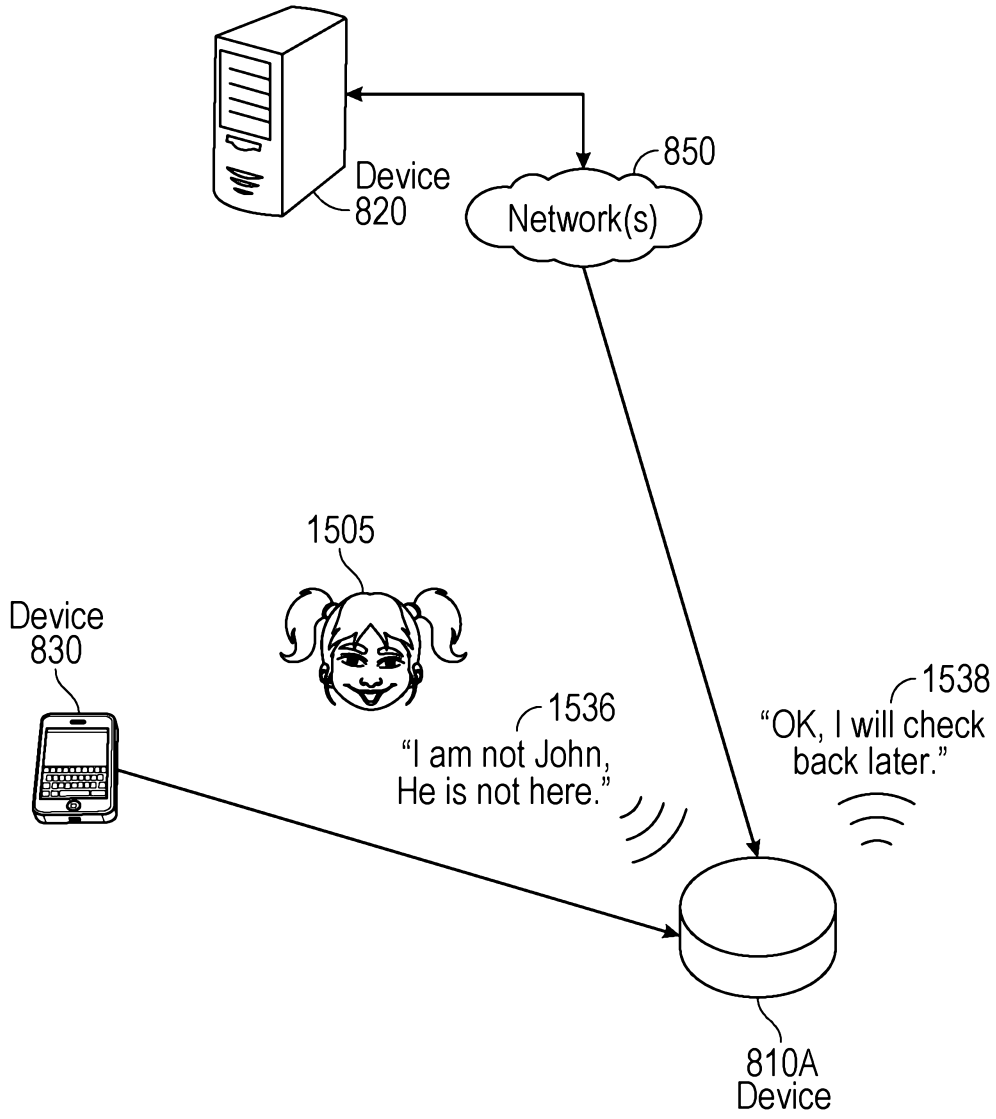

Device
820
850
Network(s)
1505
Device
830
1536
"I am not John,
He is not here."
1538
"OK, I will check
back later."
810A
Device

FIG. 15D

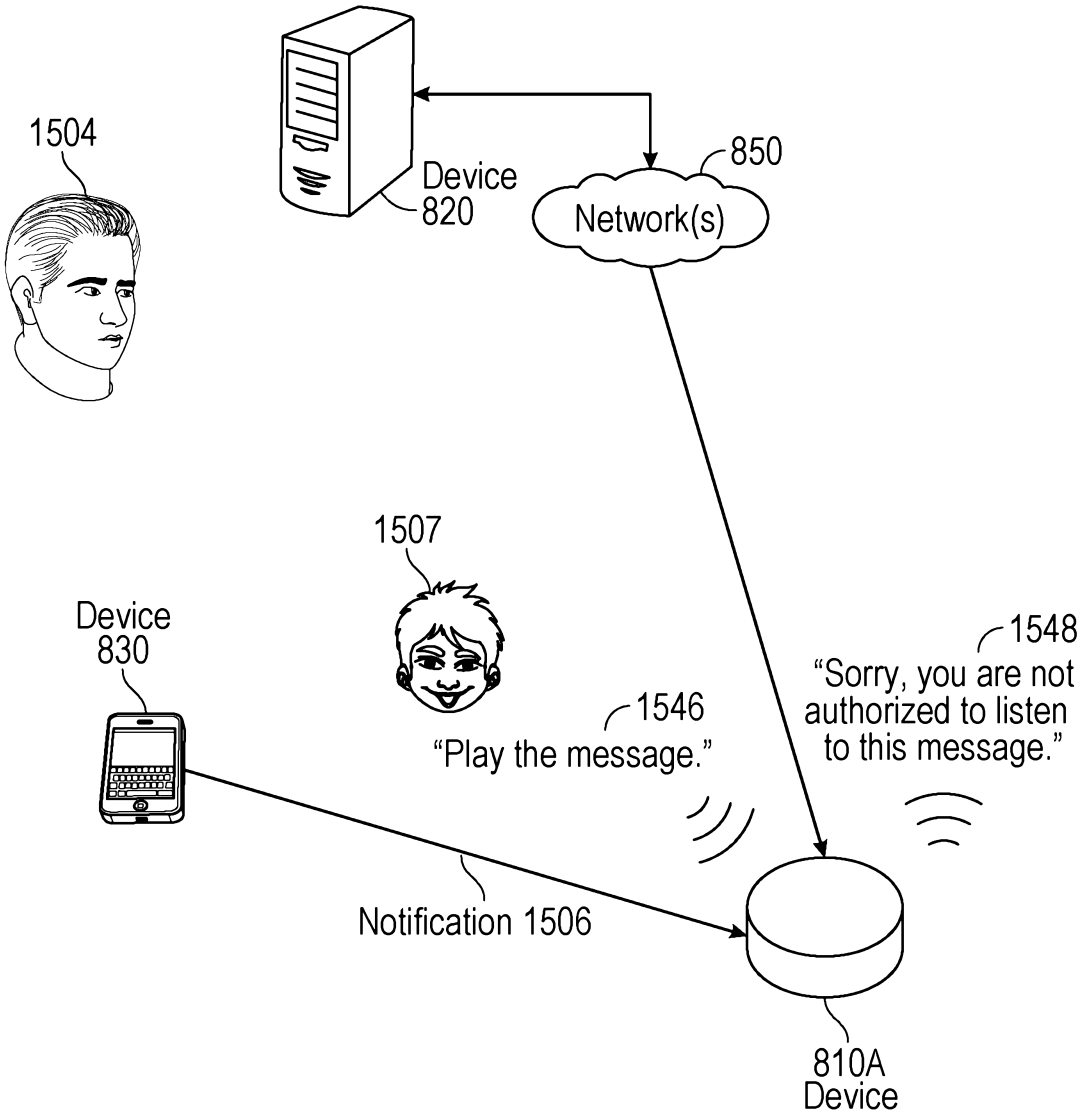

1504
Device
820
850
Network(s)
1507
Device
830
1546
"Play the message."
1548
"Sorry, you are not authorized to listen to this message."
Notification 1506
810A
Device

FIG. 15E

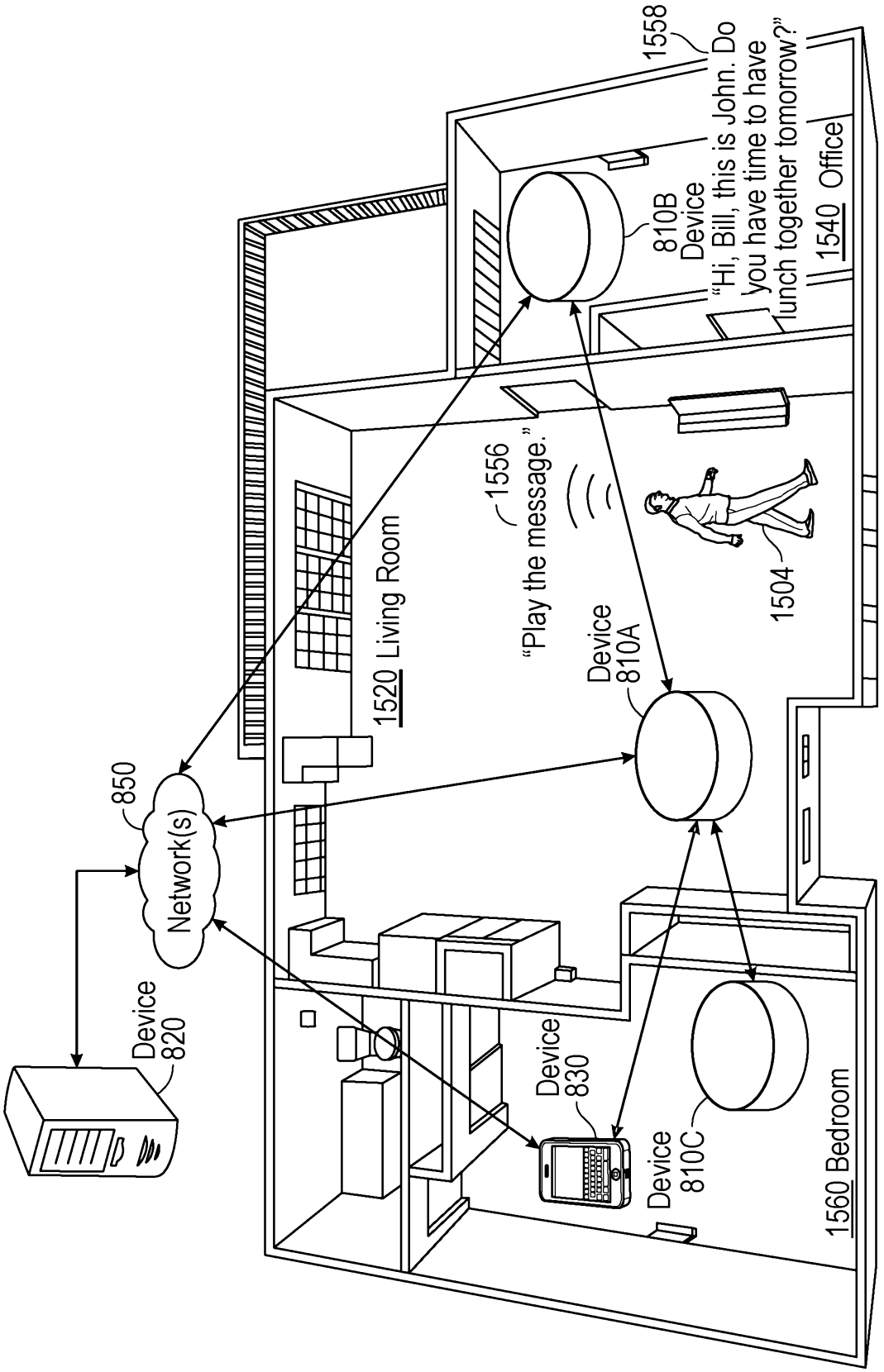

1558
"Hi, Bill, this is John. Do you have time to have lunch together tomorrow?"
810B Device
1540 Office
1556
"Play the message."
1504
1520 Living Room
Device 810A
850
Network(s)
Device 820
Device 830
Device 810C
1560 Bedroom

FIG. 15F

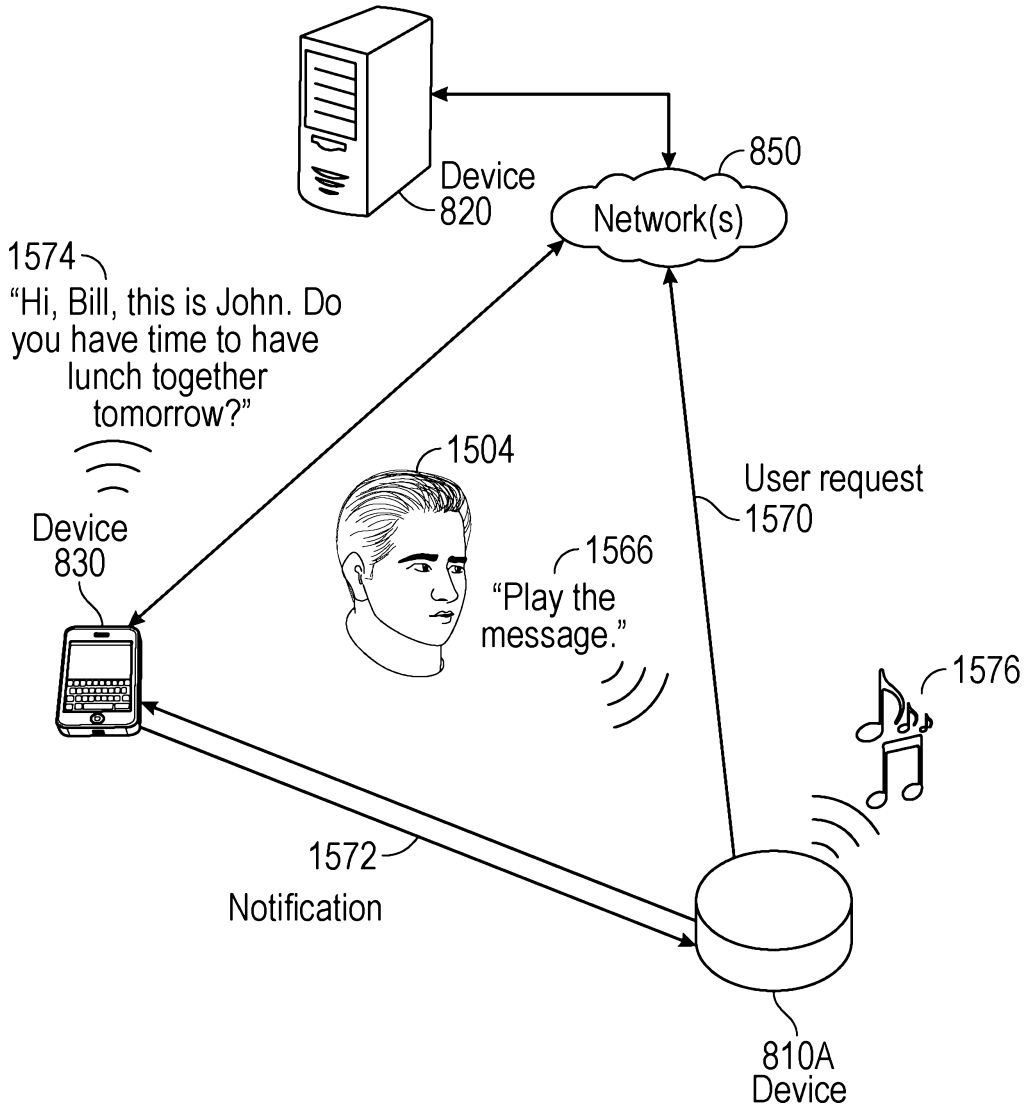

Device
820
850
Network(s)
1574
"Hi, Bill, this is John. Do you have time to have lunch together tomorrow?"
1504
User request
1570
Device
830
1566
"Play the message."
1576
1572
Notification
810A
Device

FIG. 15G

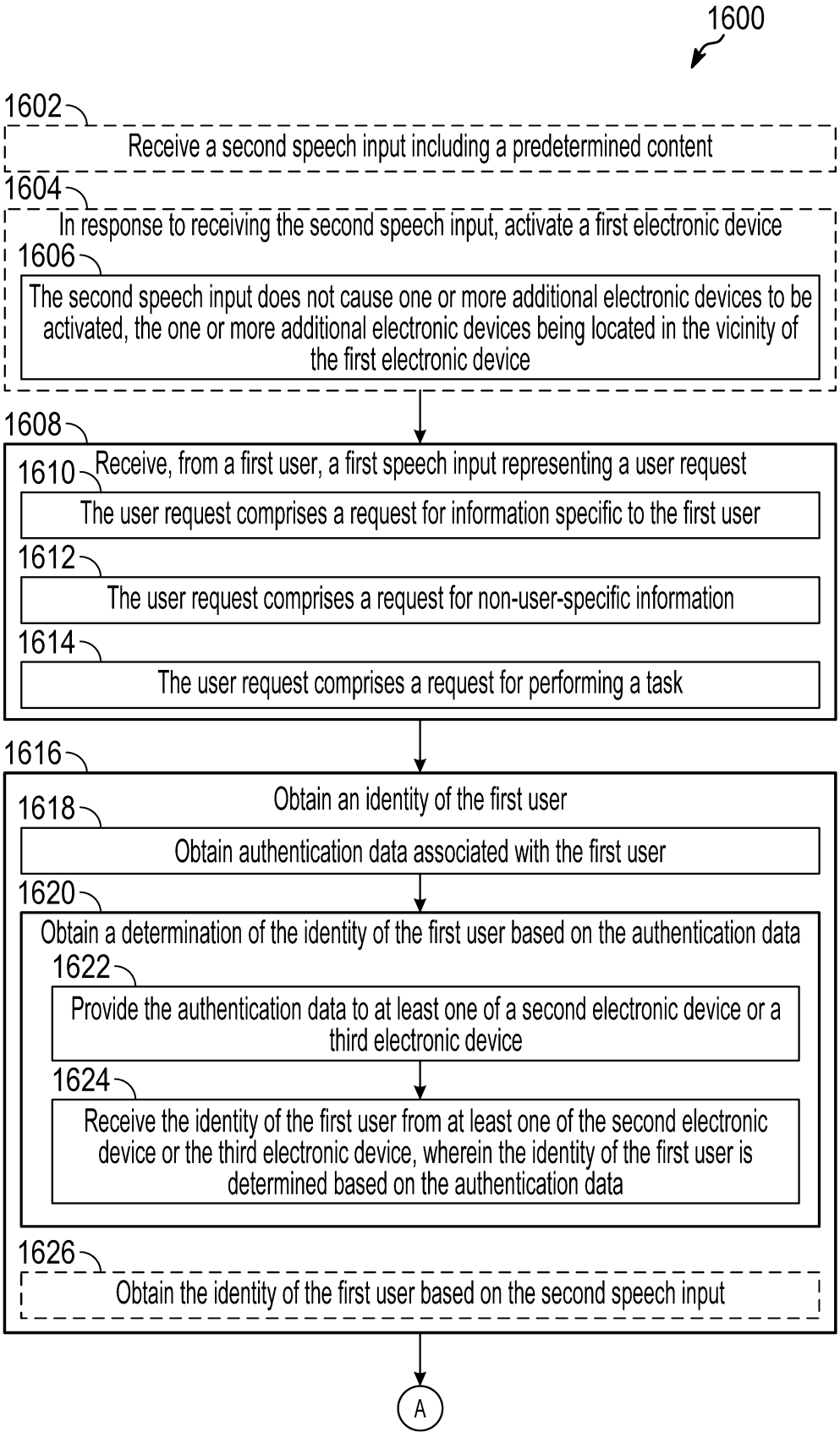

1600
1602
Receive a second speech input including a predetermined content
1604
In response to receiving the second speech input, activate a first electronic device
1606
The second speech input does not cause one or more additional electronic devices to be activated, the one or more additional electronic devices being located in the vicinity of the first electronic device
1608
1610
Receive, from a first user, a first speech input representing a user request
The user request comprises a request for information specific to the first user
1612
The user request comprises a request for non-user-specific information
1614
The user request comprises a request for performing a task
1616
Obtain an identity of the first user
1618
Obtain authentication data associated with the first user
1620
Obtain a determination of the identity of the first user based on the authentication data
1622
Provide the authentication data to at least one of a second electronic device or a third electronic device
1624
Receive the identity of the first user from at least one of the second electronic device or the third electronic device, wherein the identity of the first user is determined based on the authentication data
1626
Obtain the identity of the first user based on the second speech input
A

FIG. 16A

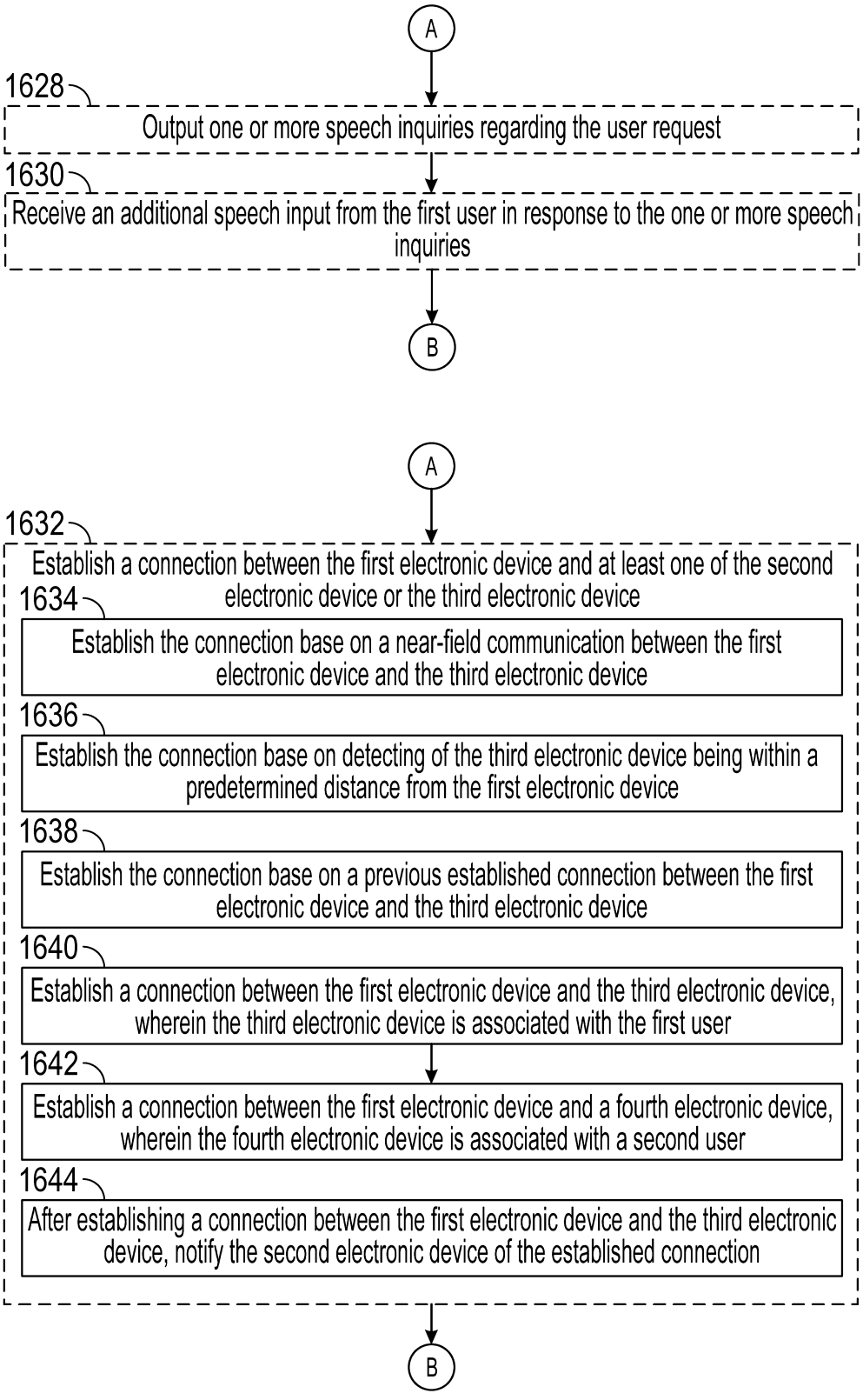

A
1628
Output one or more speech inquiries regarding the user request
1630
Receive an additional speech input from the first user in response to the one or more speech inquiries
B
A
1632
Establish a connection between the first electronic device and at least one of the second electronic device or the third electronic device
1634
Establish the connection base on a near-field communication between the first electronic device and the third electronic device
1636
Establish the connection base on detecting of the third electronic device being within a predetermined distance from the first electronic device
1638
Establish the connection base on a previous established connection between the first electronic device and the third electronic device
1640
Establish a connection between the first electronic device and the third electronic device, wherein the third electronic device is associated with the first user
1642
Establish a connection between the first electronic device and a fourth electronic device, wherein the fourth electronic device is associated with a second user
1644
After establishing a connection between the first electronic device and the third electronic device, notify the second electronic device of the established connection
B

FIG. 16B

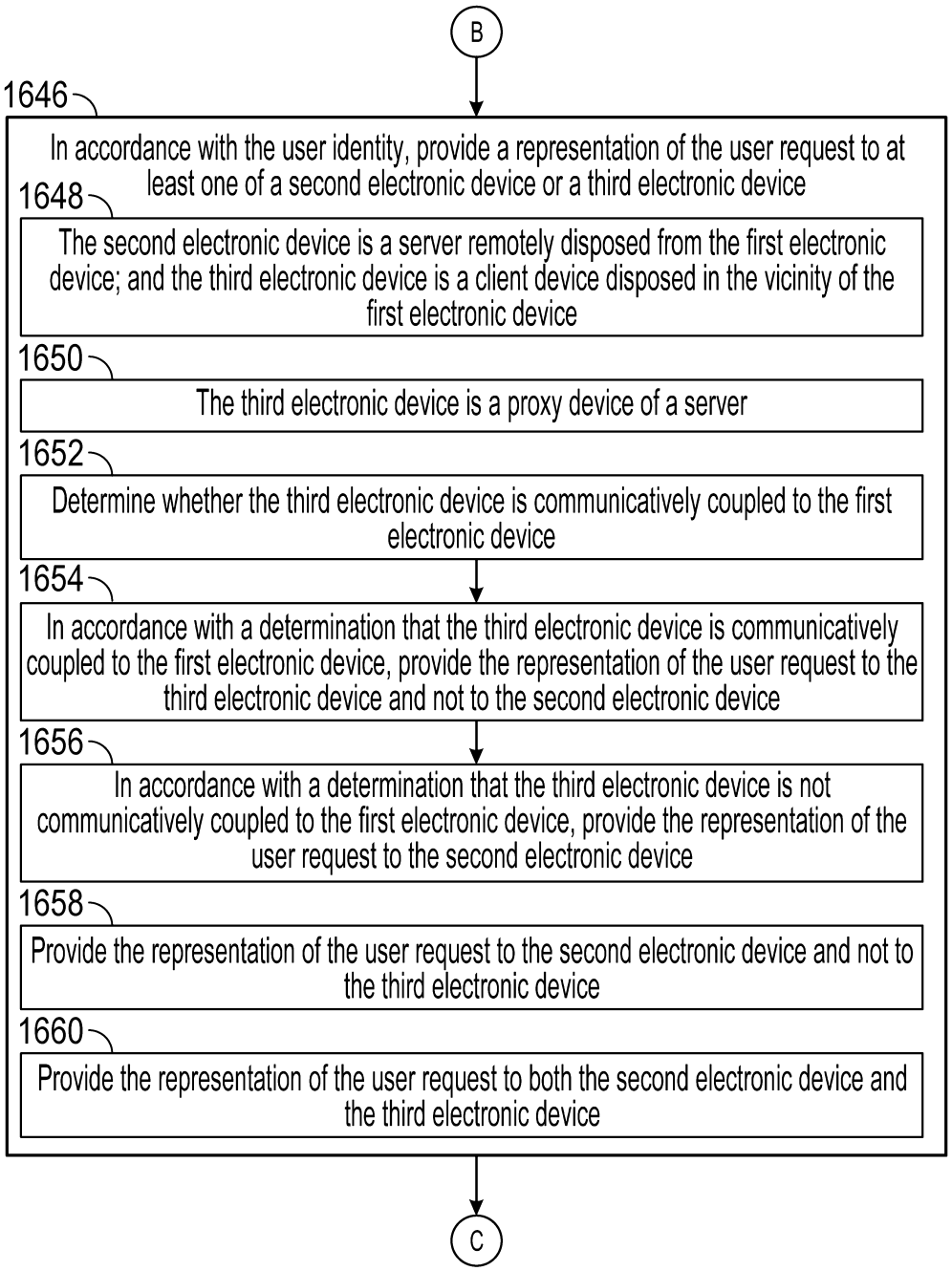

B
1646
In accordance with the user identity, provide a representation of the user request to at least one of a second electronic device or a third electronic device
1648
The second electronic device is a server remotely disposed from the first electronic device; and the third electronic device is a client device disposed in the vicinity of the first electronic device
1650
The third electronic device is a proxy device of a server
1652
Determine whether the third electronic device is communicatively coupled to the first electronic device
1654
In accordance with a determination that the third electronic device is communicatively coupled to the first electronic device, provide the representation of the user request to the third electronic device and not to the second electronic device
1656
In accordance with a determination that the third electronic device is not communicatively coupled to the first electronic device, provide the representation of the user request to the second electronic device
1658
Provide the representation of the user request to the second electronic device and not to the third electronic device
1660
Provide the representation of the user request to both the second electronic device and the third electronic device
C

FIG. 16C

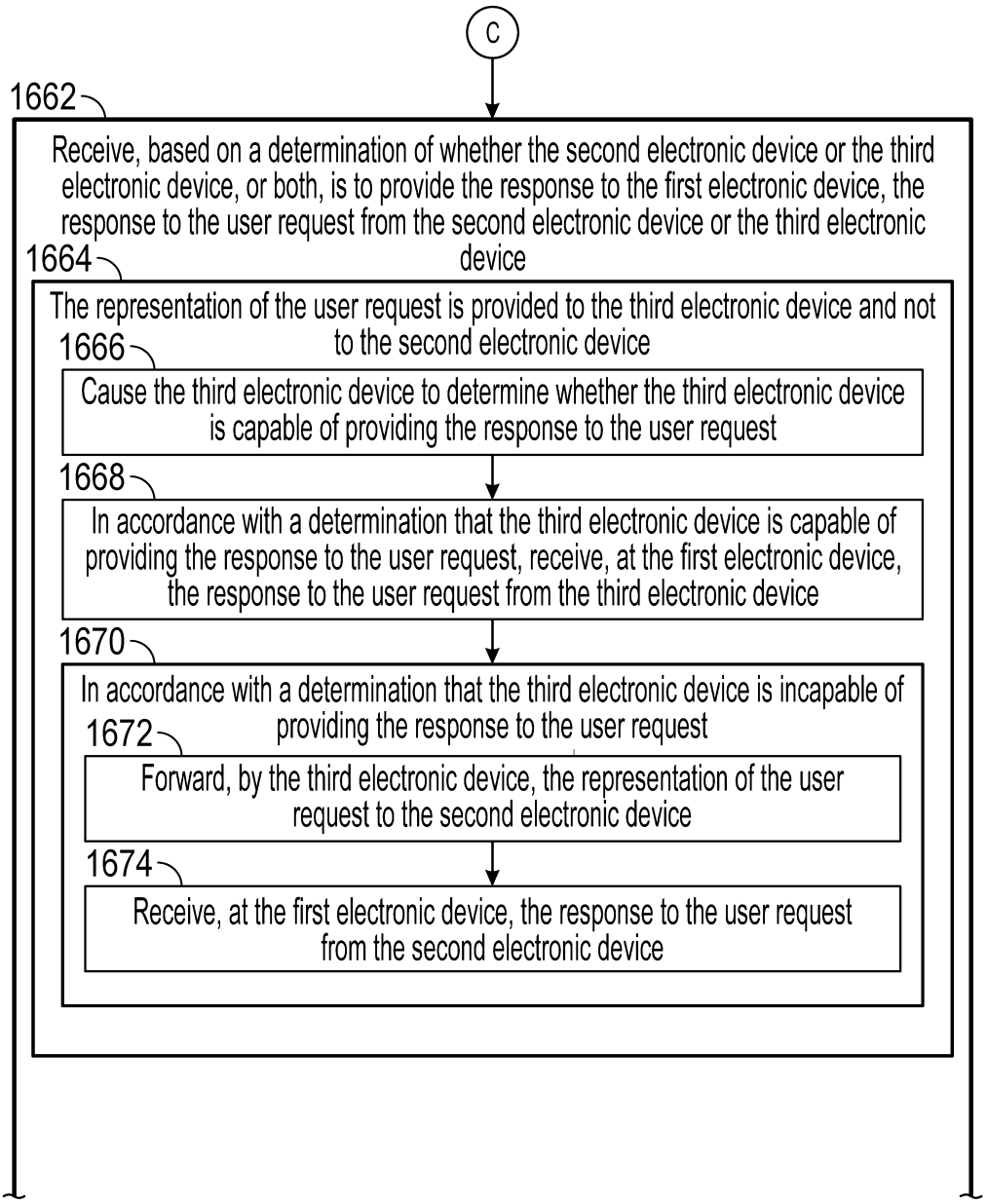

C
1662
Receive, based on a determination of whether the second electronic device or the third electronic device, or both, is to provide the response to the first electronic device, the response to the user request from the second electronic device or the third electronic device
1664
The representation of the user request is provided to the third electronic device and not to the second electronic device
1666
Cause the third electronic device to determine whether the third electronic device is capable of providing the response to the user request
1668
In accordance with a determination that the third electronic device is capable of providing the response to the user request, receive, at the first electronic device, the response to the user request from the third electronic device
1670
In accordance with a determination that the third electronic device is incapable of providing the response to the user request
1672
Forward, by the third electronic device, the representation of the user request to the second electronic device
1674
Receive, at the first electronic device, the response to the user request from the second electronic device

FIG. 16D

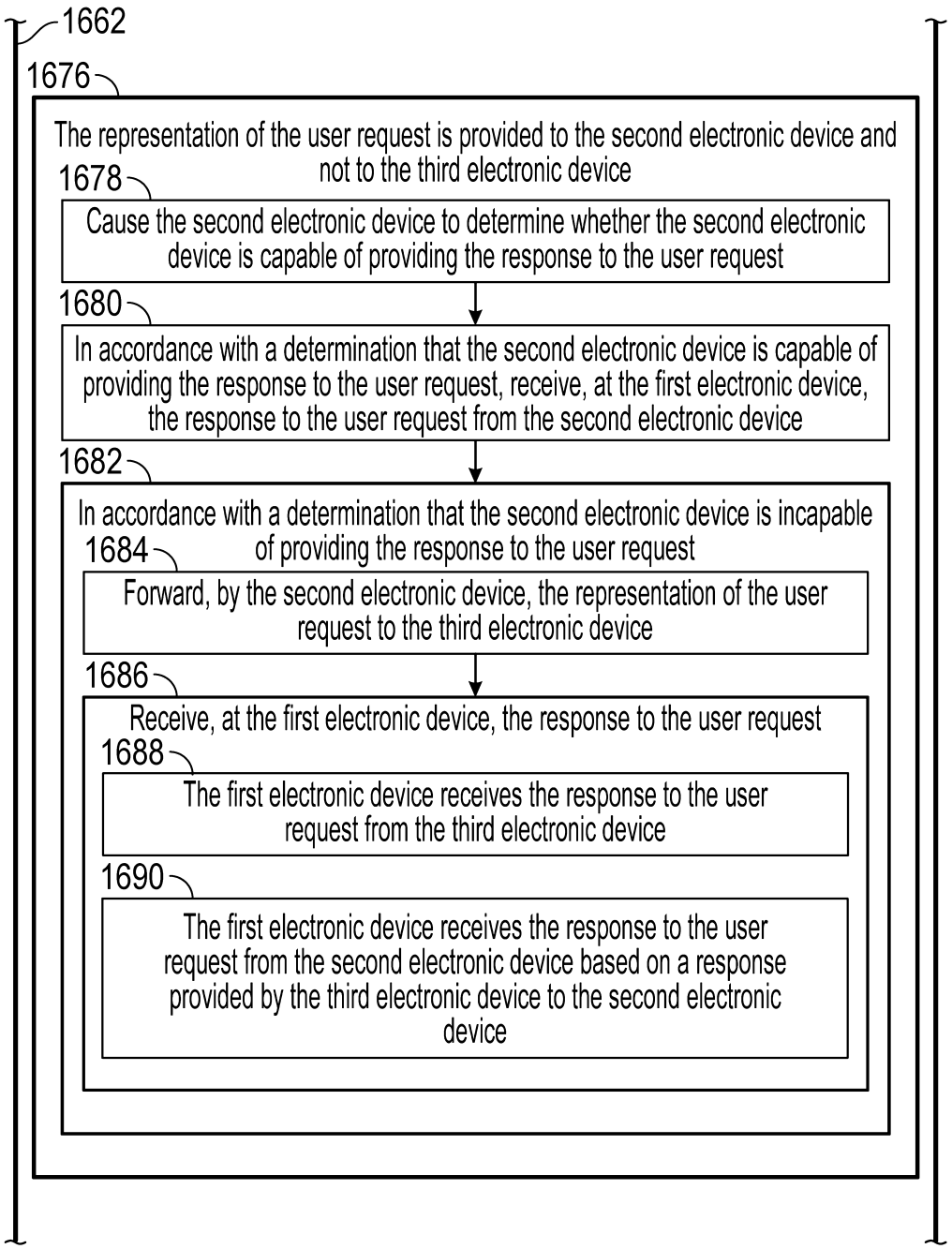

1662
1676
The representation of the user request is provided to the second electronic device and not to the third electronic device
1678
Cause the second electronic device to determine whether the second electronic device is capable of providing the response to the user request
1680
In accordance with a determination that the second electronic device is capable of providing the response to the user request, receive, at the first electronic device, the response to the user request from the second electronic device
1682
In accordance with a determination that the second electronic device is incapable of providing the response to the user request
1684
Forward, by the second electronic device, the representation of the user request to the third electronic device
1686
Receive, at the first electronic device, the response to the user request
1688
The first electronic device receives the response to the user request from the third electronic device
1690
The first electronic device receives the response to the user request from the second electronic device based on a response provided by the third electronic device to the second electronic device

FIG. 16E

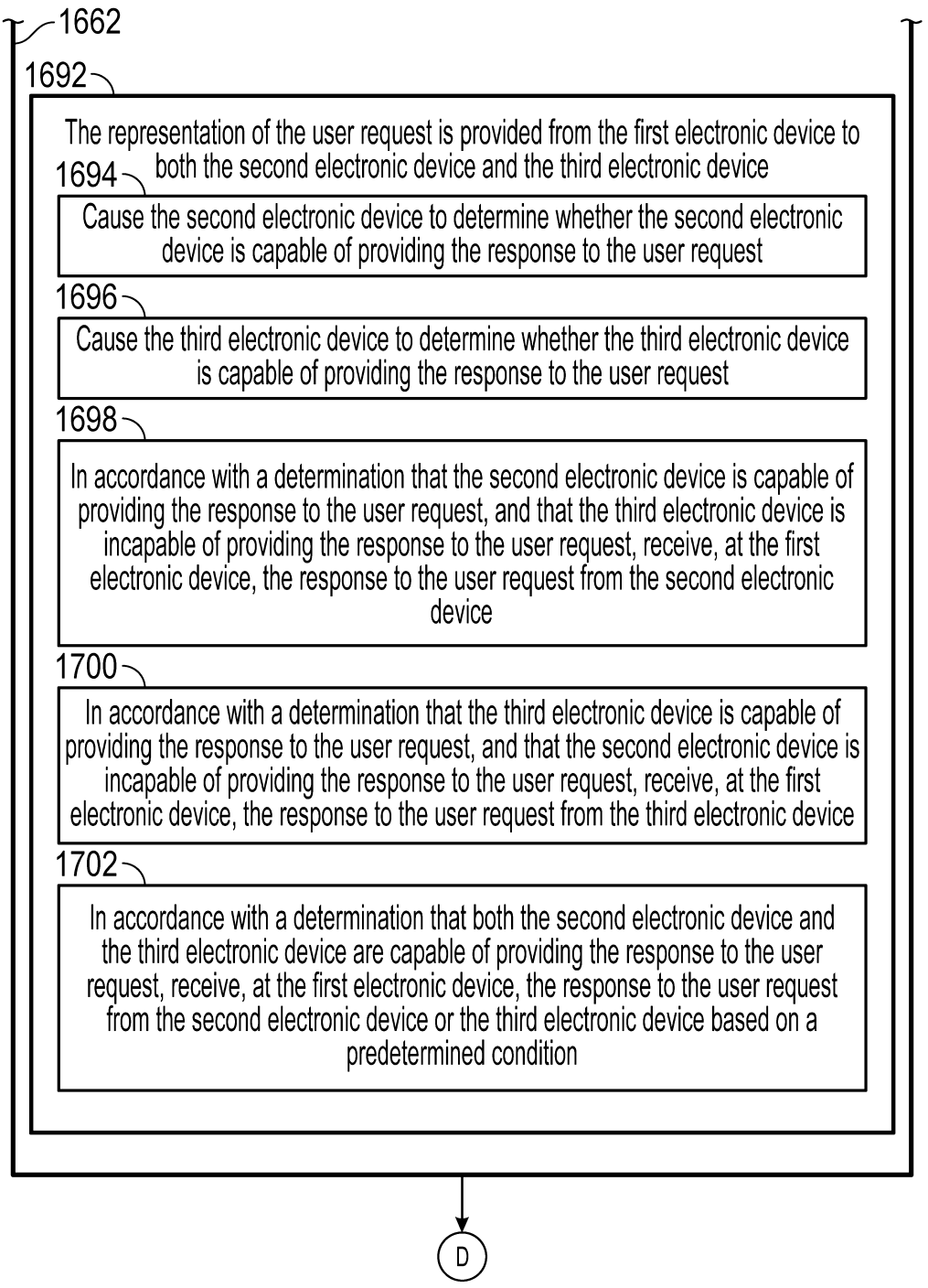

1662
1692
The representation of the user request is provided from the first electronic device to both the second electronic device and the third electronic device
1694
Cause the second electronic device to determine whether the second electronic device is capable of providing the response to the user request
1696
Cause the third electronic device to determine whether the third electronic device is capable of providing the response to the user request
1698
In accordance with a determination that the second electronic device is capable of providing the response to the user request, and that the third electronic device is incapable of providing the response to the user request, receive, at the first electronic device, the response to the user request from the second electronic device
1700
In accordance with a determination that the third electronic device is capable of providing the response to the user request, and that the second electronic device is incapable of providing the response to the user request, receive, at the first electronic device, the response to the user request from the third electronic device
1702
In accordance with a determination that both the second electronic device and the third electronic device are capable of providing the response to the user request, receive, at the first electronic device, the response to the user request from the second electronic device or the third electronic device based on a predetermined condition
D

FIG. 16F

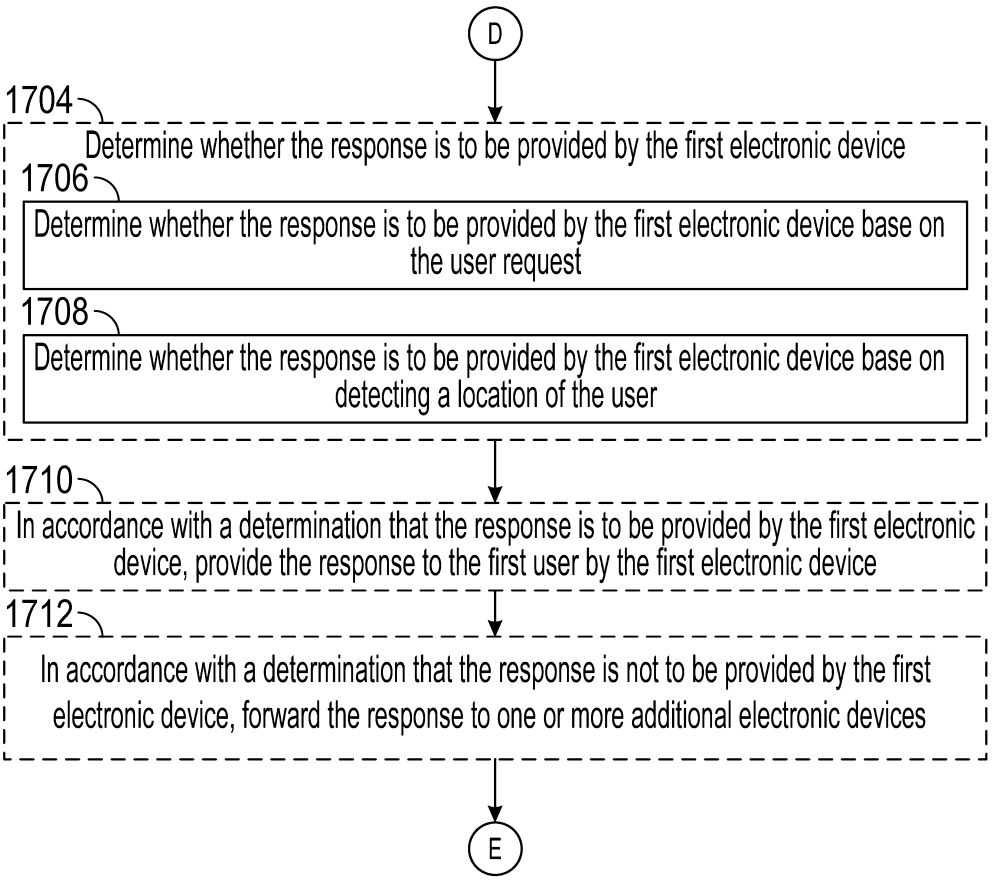

D
1704
Determine whether the response is to be provided by the first electronic device
1706
Determine whether the response is to be provided by the first electronic device base on the user request
1708
Determine whether the response is to be provided by the first electronic device base on detecting a location of the user
1710
In accordance with a determination that the response is to be provided by the first electronic device, provide the response to the first user by the first electronic device
1712
In accordance with a determination that the response is not to be provided by the first electronic device, forward the response to one or more additional electronic devices
E

FIG. 16G

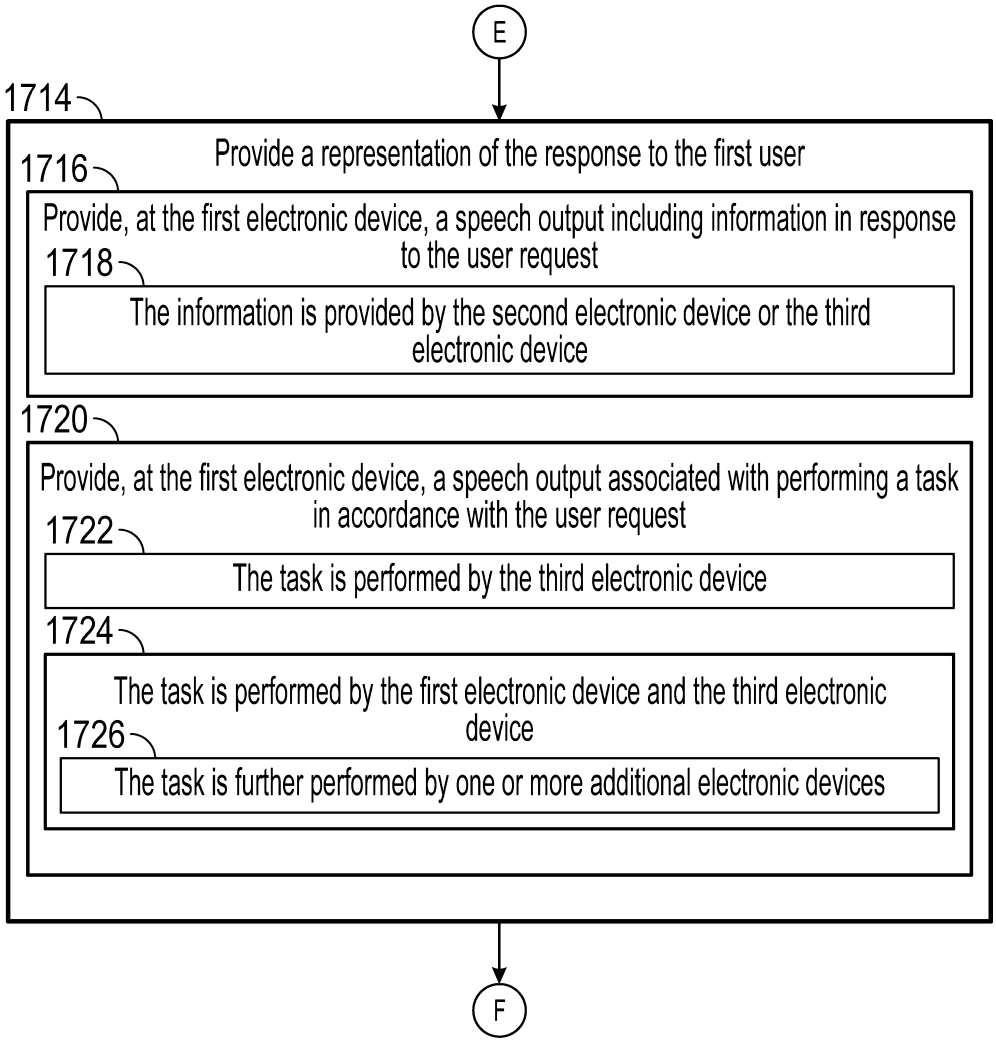

E
1714
Provide a representation of the response to the first user
1716
Provide, at the first electronic device, a speech output including information in response to the user request
1718
The information is provided by the second electronic device or the third electronic device
1720
Provide, at the first electronic device, a speech output associated with performing a task in accordance with the user request
1722
The task is performed by the third electronic device
1724
The task is performed by the first electronic device and the third electronic device
1726
The task is further performed by one or more additional electronic devices
F

FIG. 16H

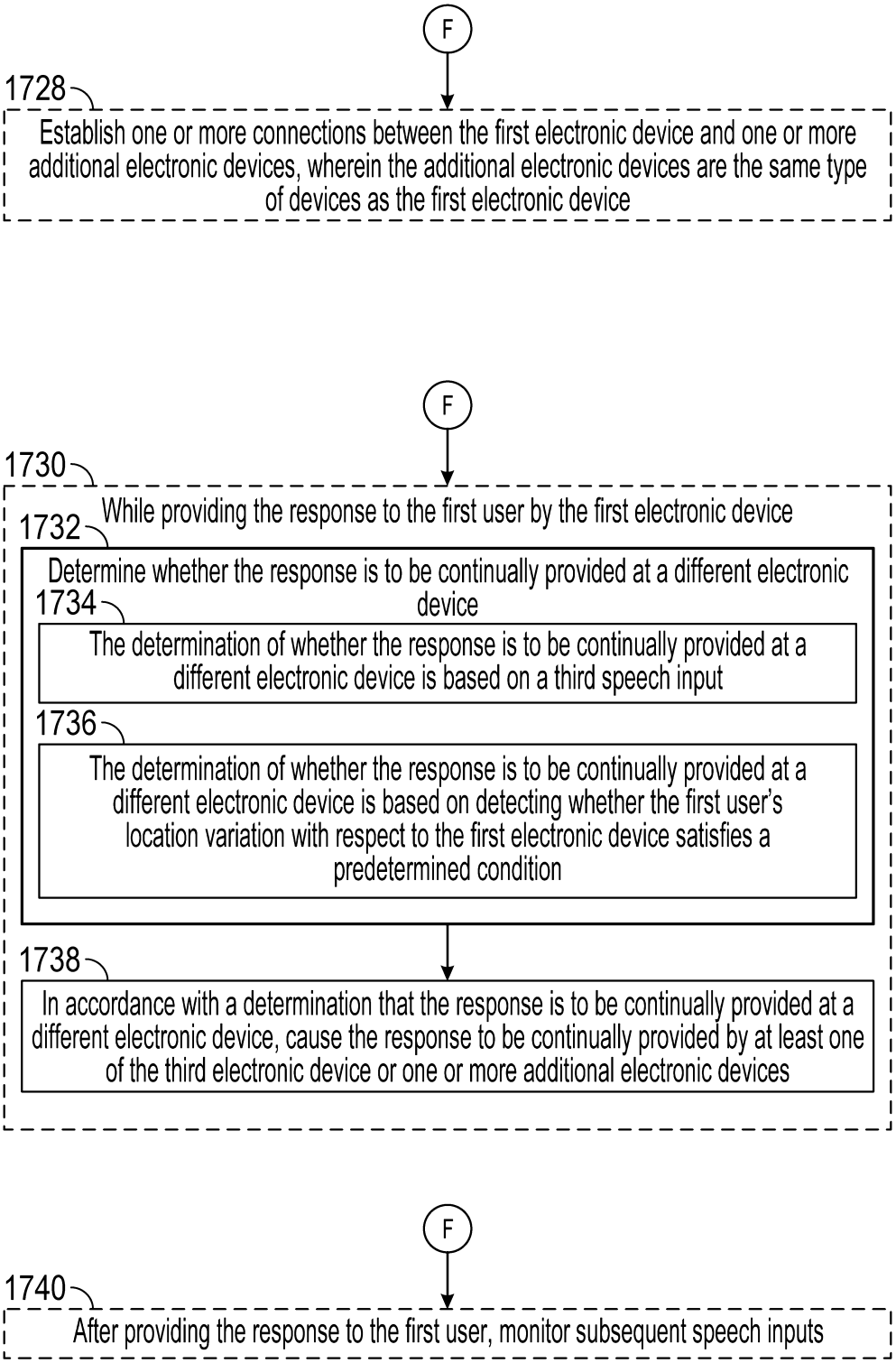

F
1728
Establish one or more connections between the first electronic device and one or more additional electronic devices, wherein the additional electronic devices are the same type of devices as the first electronic device
F
1730
While providing the response to the first user by the first electronic device
1732
Determine whether the response is to be continually provided at a different electronic device
1734
The determination of whether the response is to be continually provided at a different electronic device is based on a third speech input
1736
The determination of whether the response is to be continually provided at a different electronic device is based on detecting whether the first user's location variation with respect to the first electronic device satisfies a predetermined condition
1738
In accordance with a determination that the response is to be continually provided at a different electronic device, cause the response to be continually provided by at least one of the third electronic device or one or more additional electronic devices
F
1740
After providing the response to the first user, monitor subsequent speech inputs

FIG. 16I

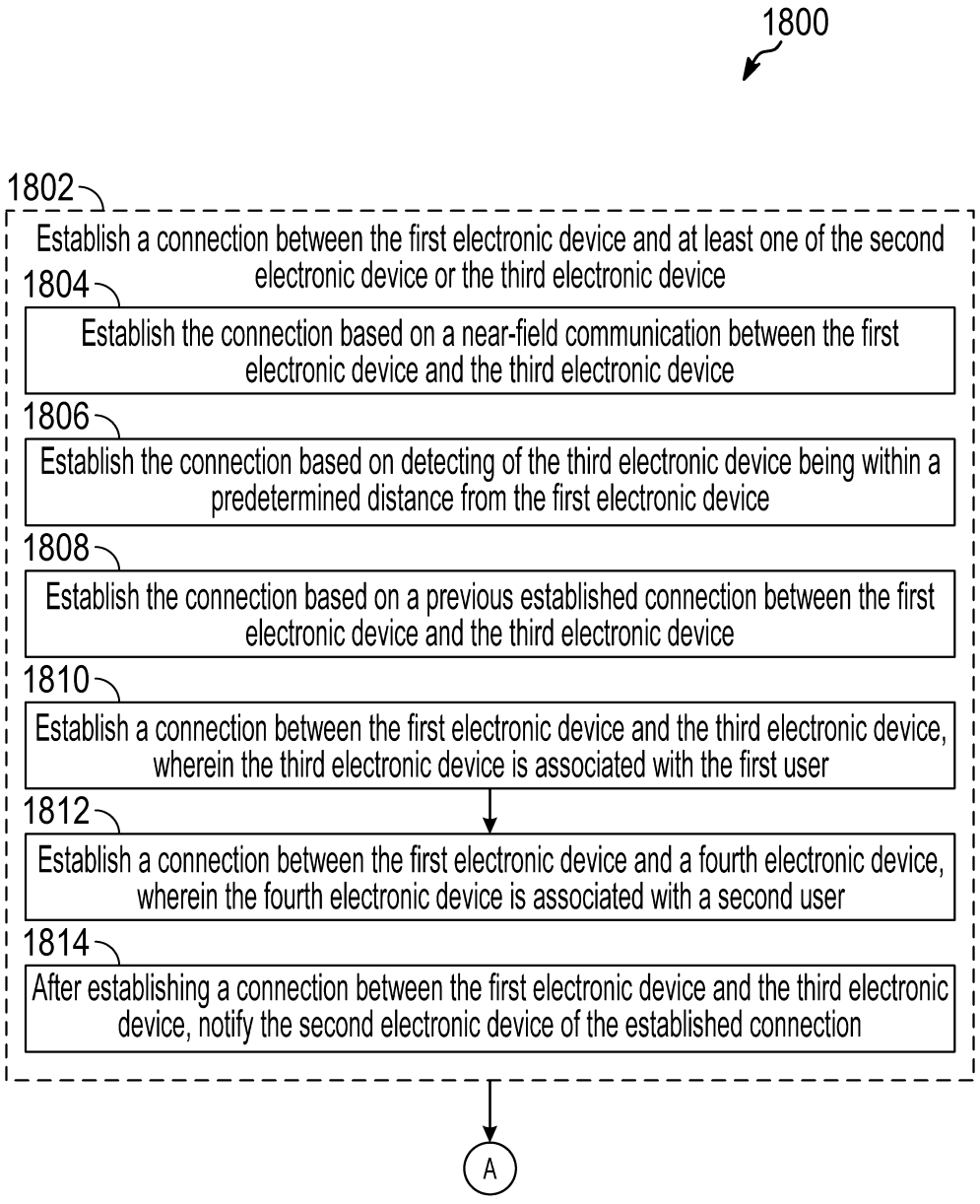

1800
1802
Establish a connection between the first electronic device and at least one of the second electronic device or the third electronic device
1804
Establish the connection based on a near-field communication between the first electronic device and the third electronic device
1806
Establish the connection based on detecting of the third electronic device being within a predetermined distance from the first electronic device
1808
Establish the connection based on a previous established connection between the first electronic device and the third electronic device
1810
Establish a connection between the first electronic device and the third electronic device, wherein the third electronic device is associated with the first user
1812
Establish a connection between the first electronic device and a fourth electronic device, wherein the fourth electronic device is associated with a second user
1814
After establishing a connection between the first electronic device and the third electronic device, notify the second electronic device of the established connection
A

FIG. 17A

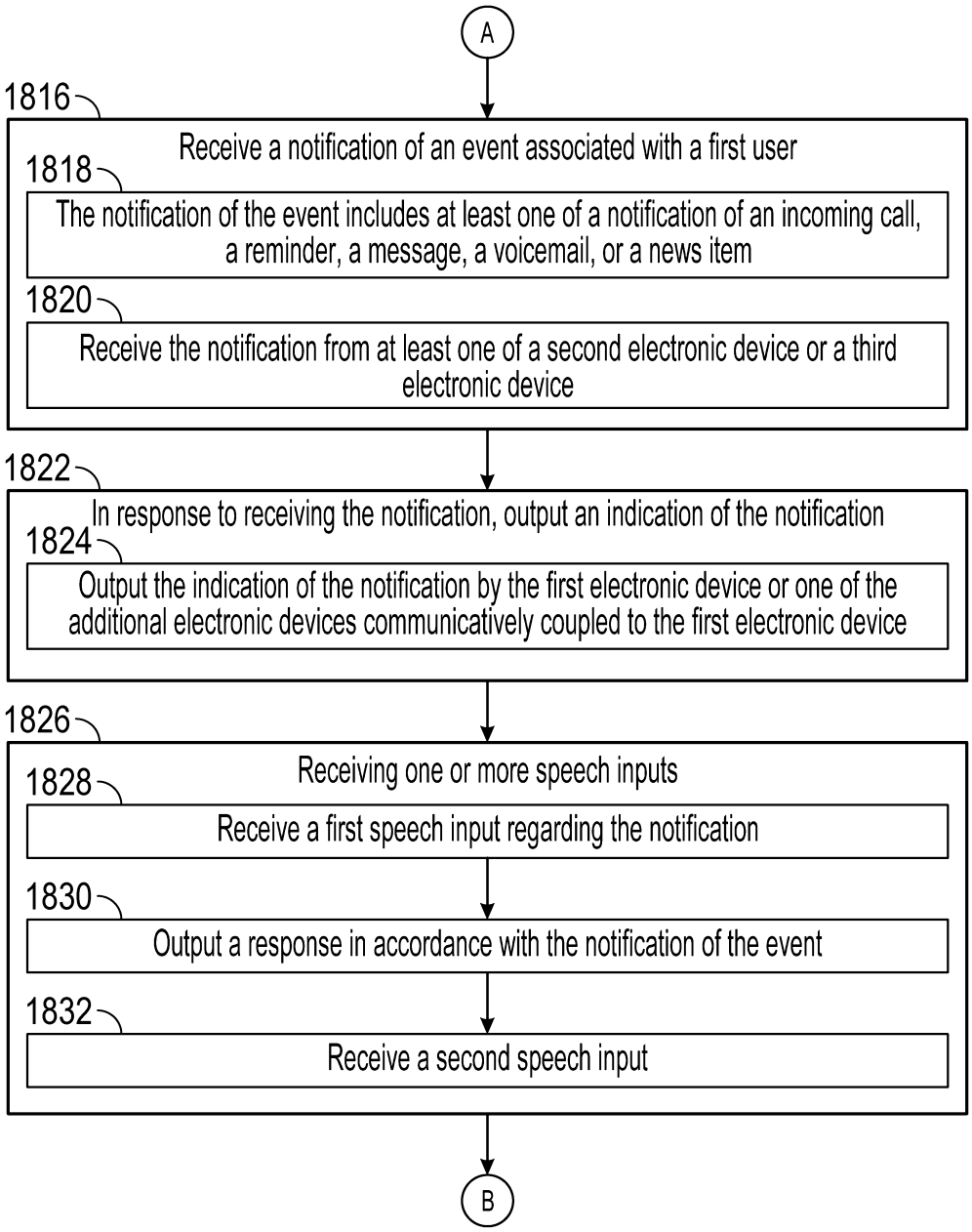

A
1816
Receive a notification of an event associated with a first user
1818
The notification of the event includes at least one of a notification of an incoming call, a reminder, a message, a voicemail, or a news item
1820
Receive the notification from at least one of a second electronic device or a third electronic device
1822
In response to receiving the notification, output an indication of the notification
1824
Output the indication of the notification by the first electronic device or one of the additional electronic devices communicatively coupled to the first electronic device
1826
Receiving one or more speech inputs
1828
Receive a first speech input regarding the notification
1830
Output a response in accordance with the notification of the event
1832
Receive a second speech input
B

FIG. 17B

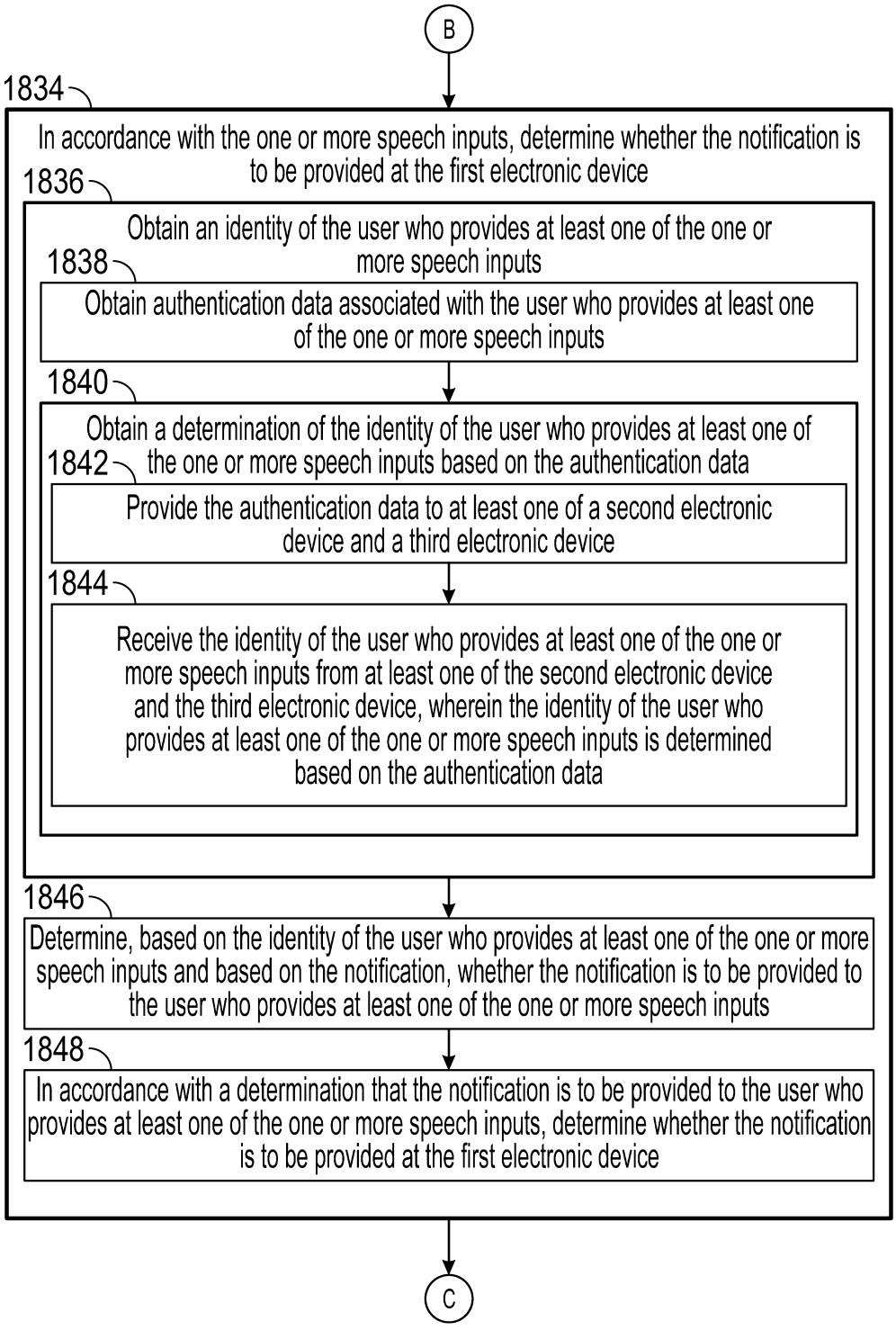

B
1834
In accordance with the one or more speech inputs, determine whether the notification is to be provided at the first electronic device
1836
Obtain an identity of the user who provides at least one of the one or more speech inputs
1838
Obtain authentication data associated with the user who provides at least one of the one or more speech inputs
1840
Obtain a determination of the identity of the user who provides at least one of the one or more speech inputs based on the authentication data
1842
Provide the authentication data to at least one of a second electronic device and a third electronic device
1844
Receive the identity of the user who provides at least one of the one or more speech inputs from at least one of the second electronic device and the third electronic device, wherein the identity of the user who provides at least one of the one or more speech inputs is determined based on the authentication data
1846
Determine, based on the identity of the user who provides at least one of the one or more speech inputs and based on the notification, whether the notification is to be provided to the user who provides at least one of the one or more speech inputs
1848
In accordance with a determination that the notification is to be provided to the user who provides at least one of the one or more speech inputs, determine whether the notification is to be provided at the first electronic device
C

FIG. 17C

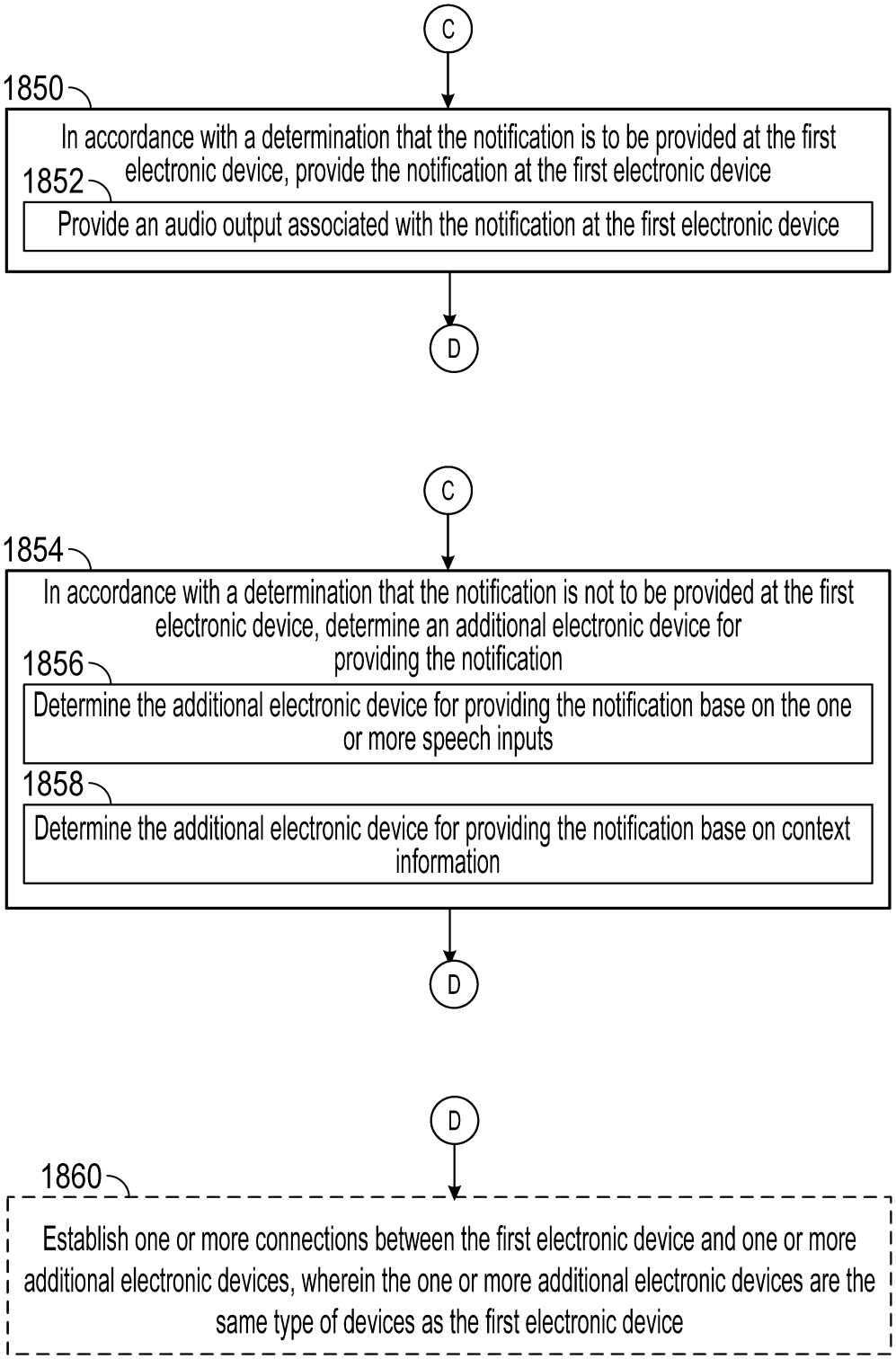

C
1850
In accordance with a determination that the notification is to be provided at the first electronic device, provide the notification at the first electronic device
1852
Provide an audio output associated with the notification at the first electronic device
D
C
1854
In accordance with a determination that the notification is not to be provided at the first electronic device, determine an additional electronic device for providing the notification
1856
Determine the additional electronic device for providing the notification base on the one or more speech inputs
1858
Determine the additional electronic device for providing the notification base on context information
D
D
1860
Establish one or more connections between the first electronic device and one or more additional electronic devices, wherein the one or more additional electronic devices are the same type of devices as the first electronic device

FIG. 17D

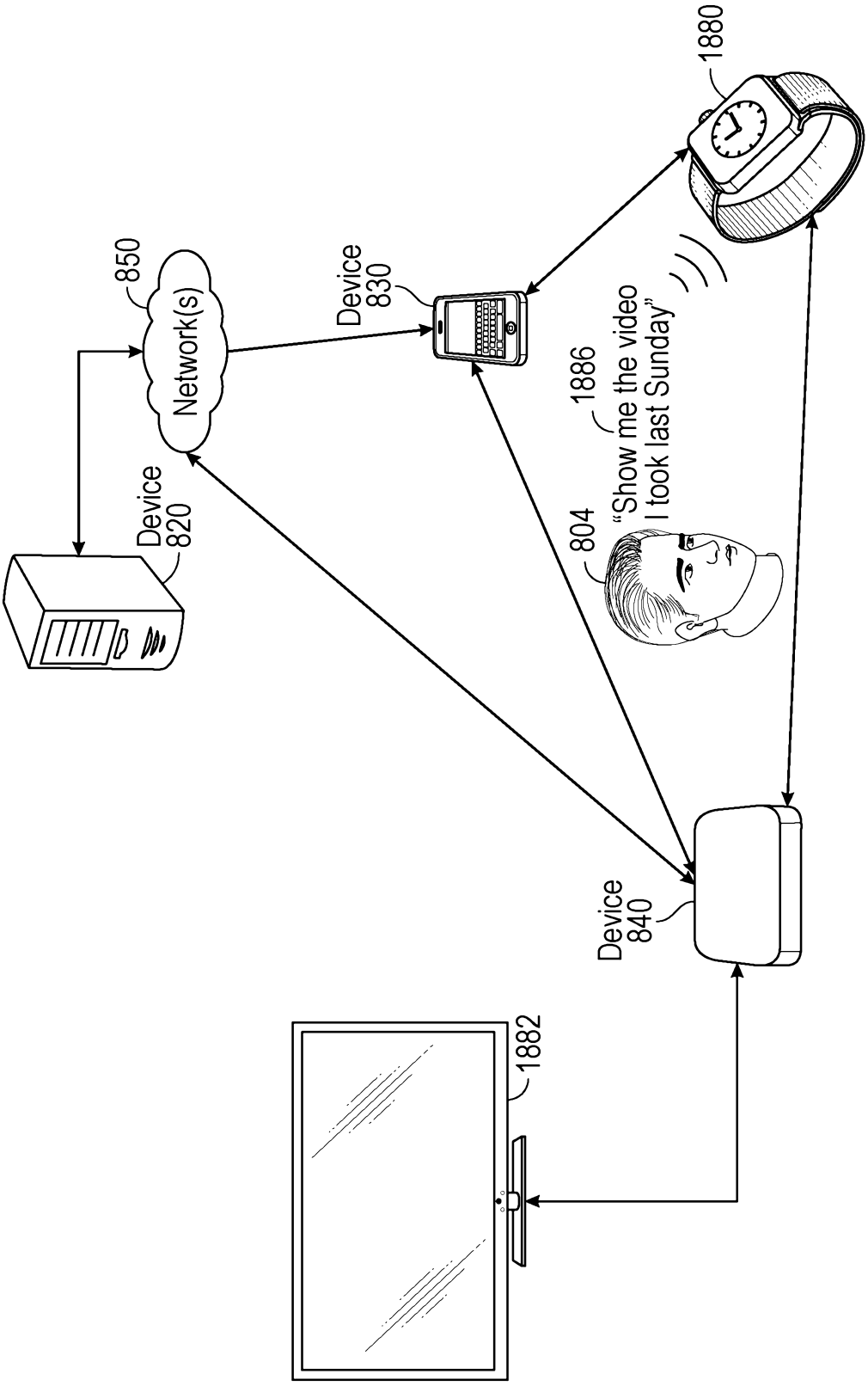

1880
850
Network(s)
Device
830
1886
"Show me the video
I took last Sunday"
804
Device
820
Device
840
1882

FIG. 18A

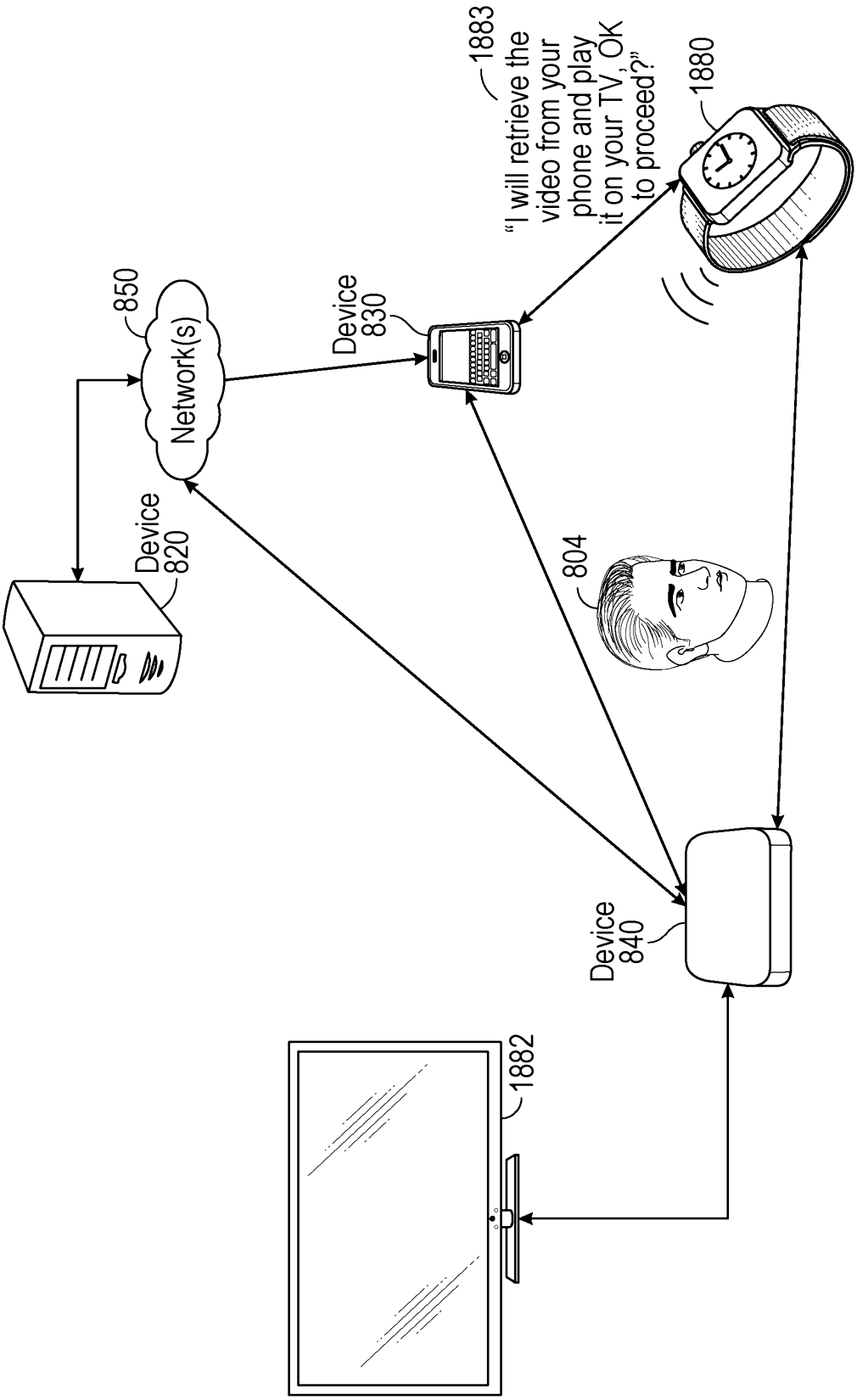

1883
"I will retrieve the video from your phone and play it on your TV, OK to proceed?"
1880
850
Network(s)
Device 830
Device 820
804
Device 840
1882

FIG. 18B

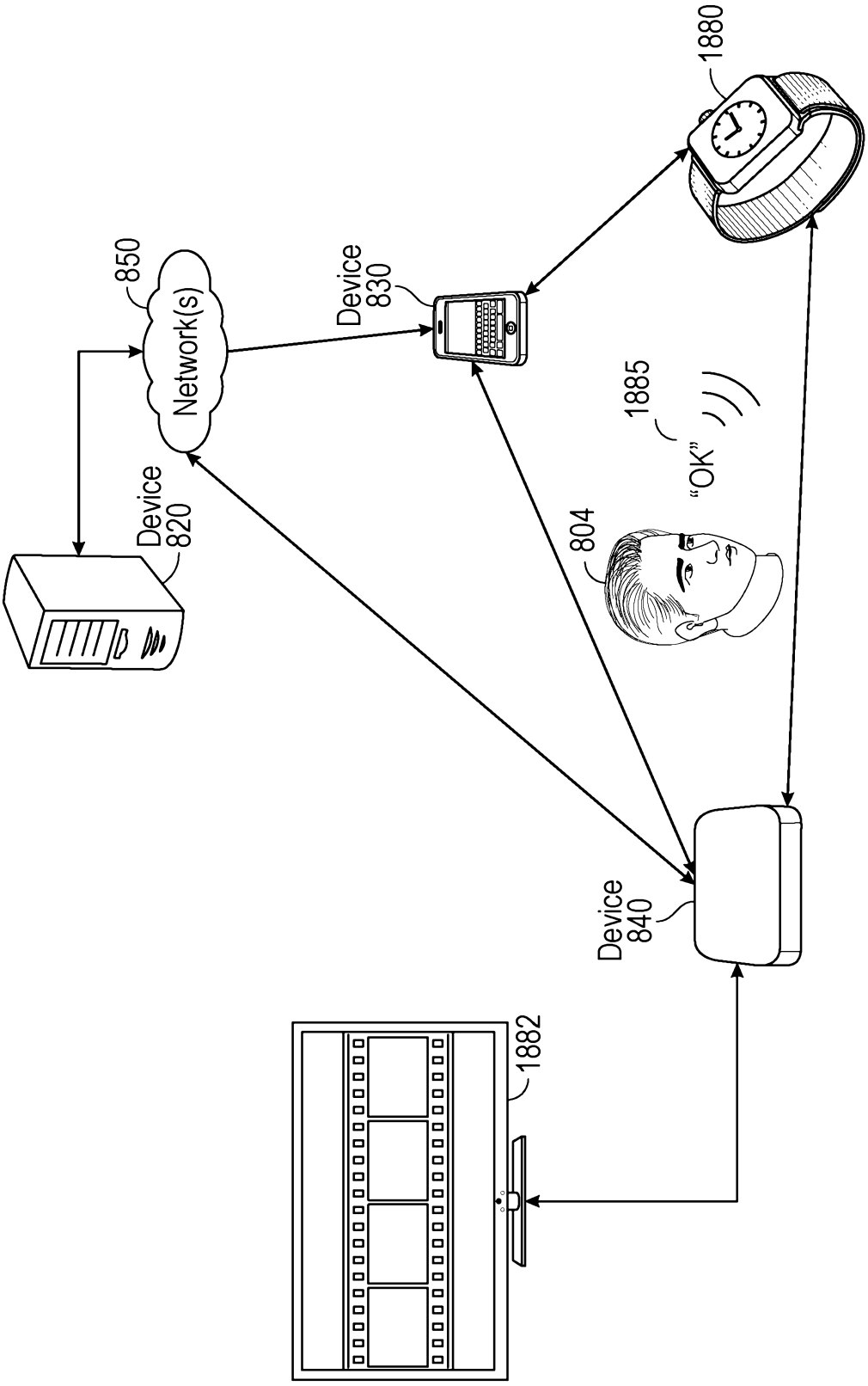

1880
850
Network(s)
Device
830
1885
"OK"
804
Device
820
Device
840
1882

FIG. 18C

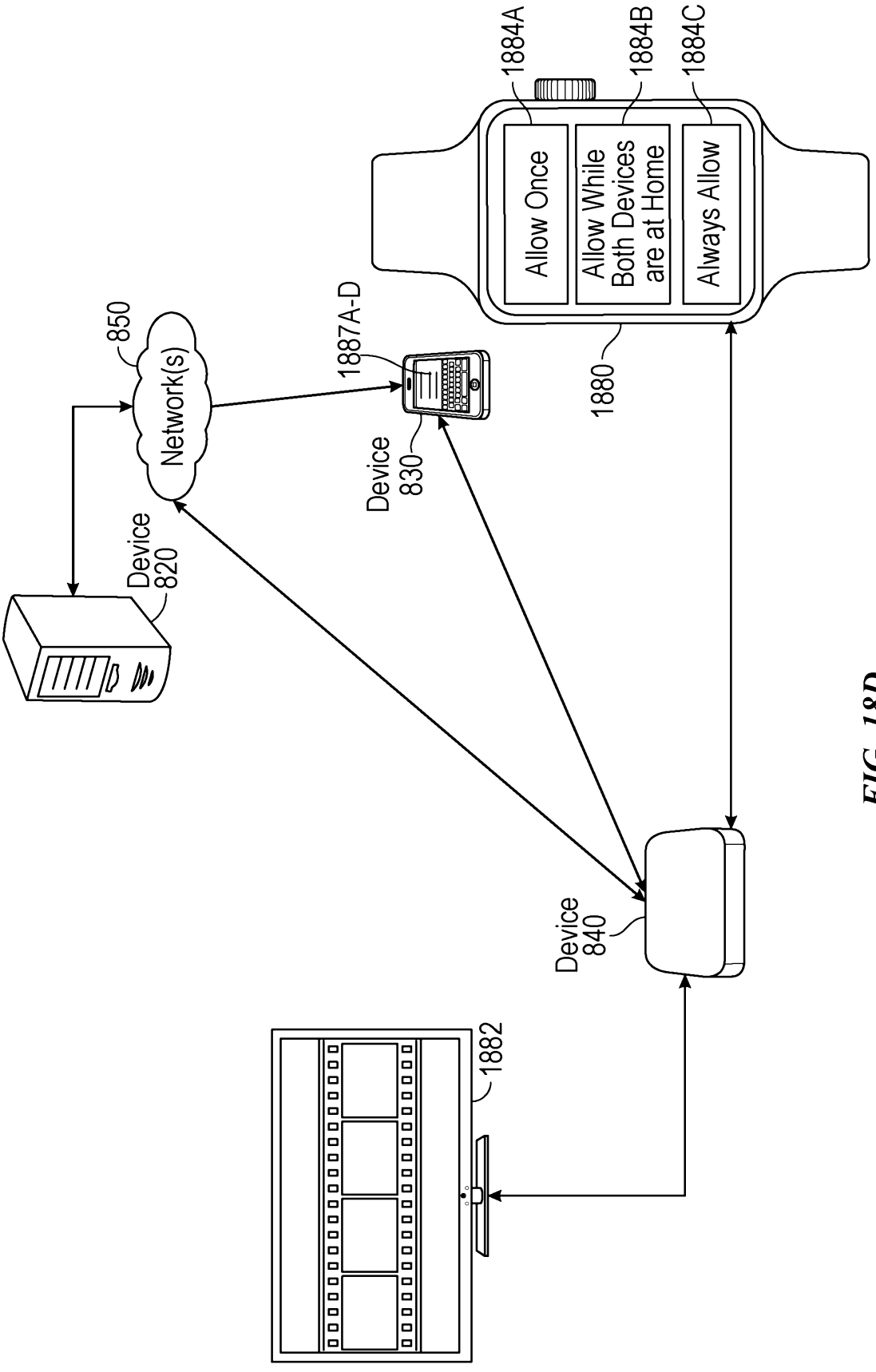

1884A
1884B
1884C
Allow Once
Allow While Both Devices are at Home
Always Allow
1880
850
Network(s)
1887A-D
Device 830
Device 820
Device 840
1882

FIG. 18D

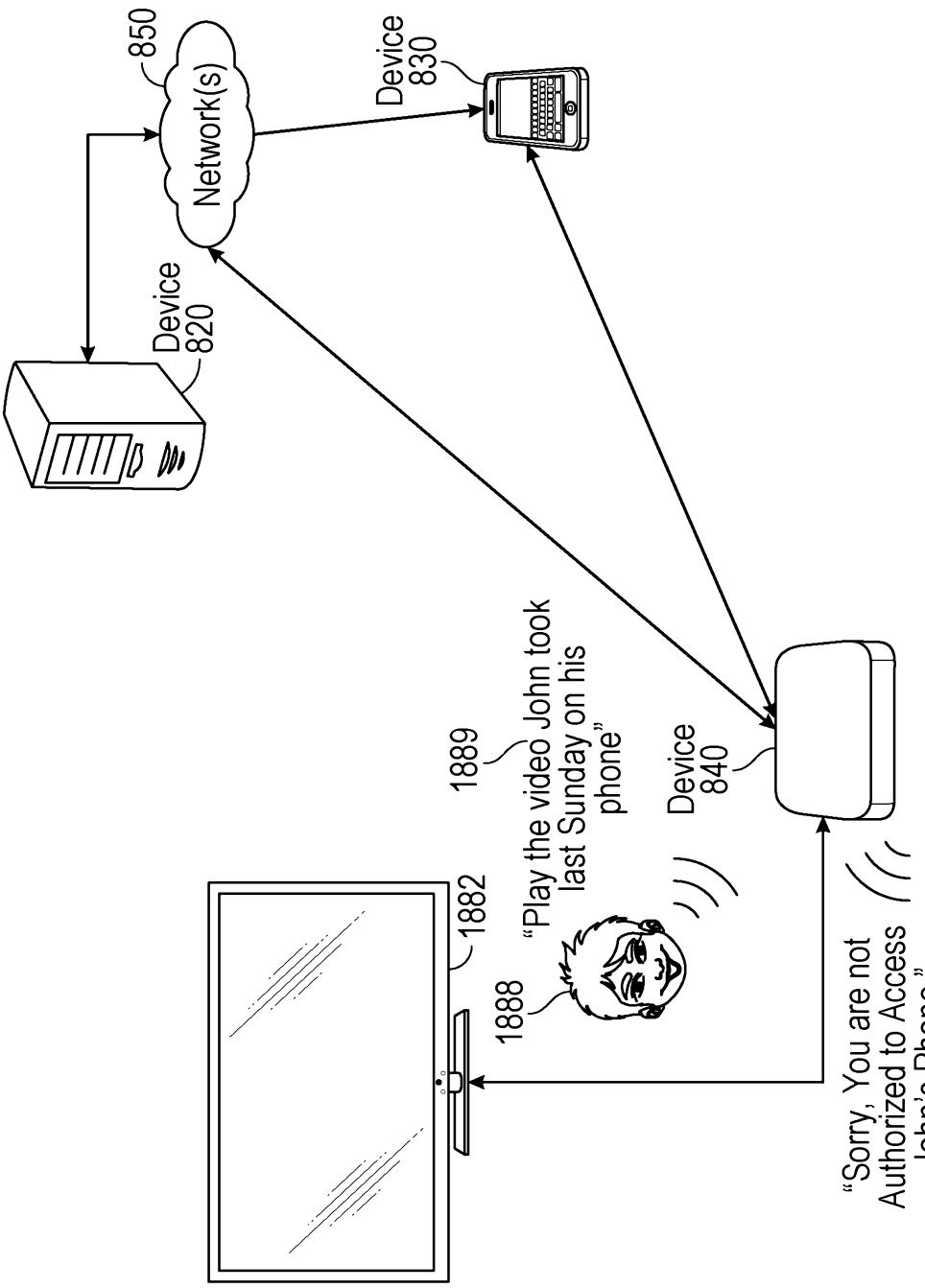

850
Network(s)
Device
830
Device
820
1889
"Play the video John took last Sunday on his phone"
Device
840
1882
1888
"Sorry, You are not Authorized to Access John's Phone."

FIG. 18E

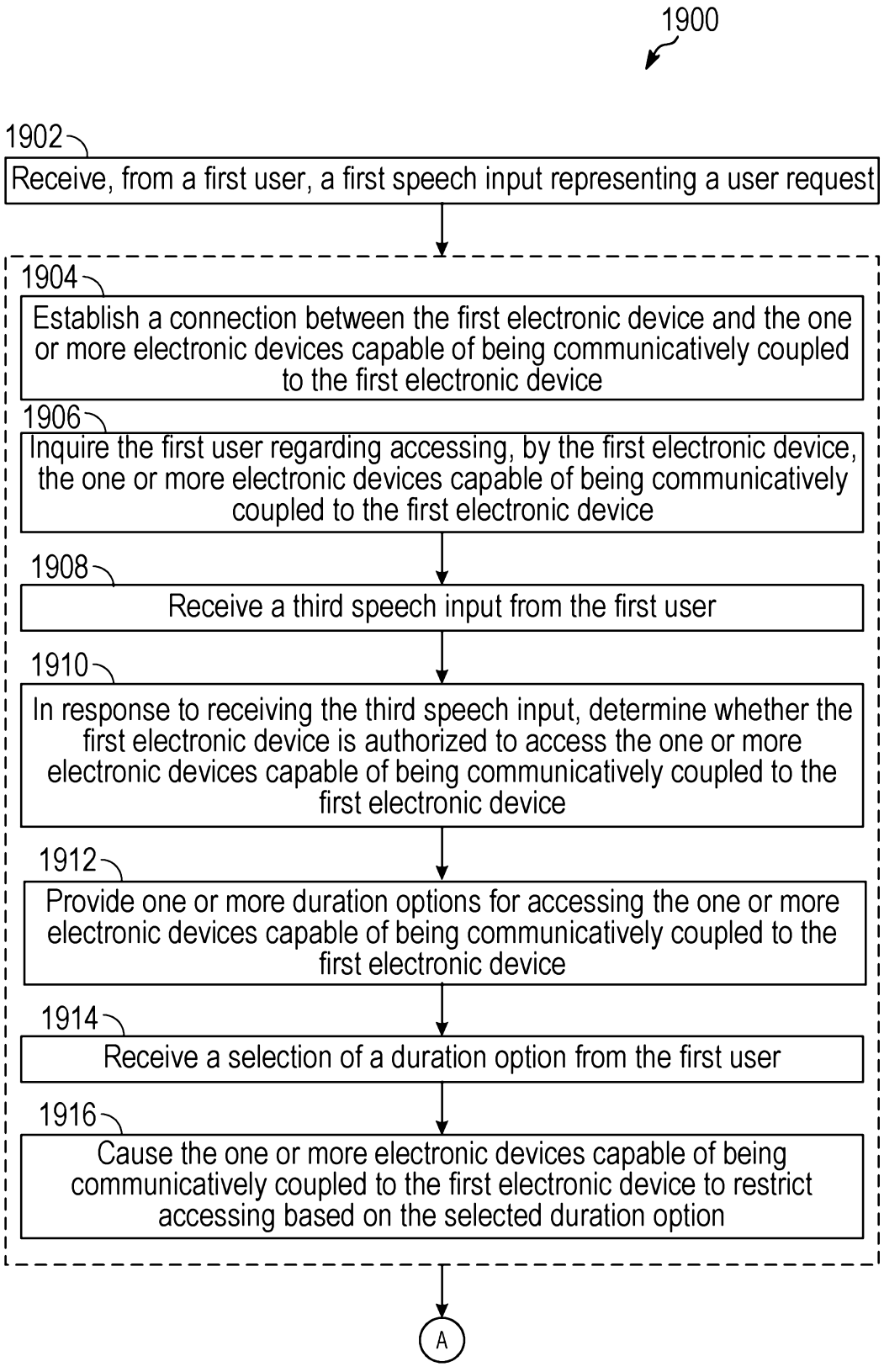

1900
1902
Receive, from a first user, a first speech input representing a user request
1904
Establish a connection between the first electronic device and the one or more electronic devices capable of being communicatively coupled to the first electronic device
1906
Inquire the first user regarding accessing, by the first electronic device, the one or more electronic devices capable of being communicatively coupled to the first electronic device
1908
Receive a third speech input from the first user
1910
In response to receiving the third speech input, determine whether the first electronic device is authorized to access the one or more electronic devices capable of being communicatively coupled to the first electronic device
1912
Provide one or more duration options for accessing the one or more electronic devices capable of being communicatively coupled to the first electronic device
1914
Receive a selection of a duration option from the first user
1916
Cause the one or more electronic devices capable of being communicatively coupled to the first electronic device to restrict accessing based on the selected duration option
A

FIG. 19A

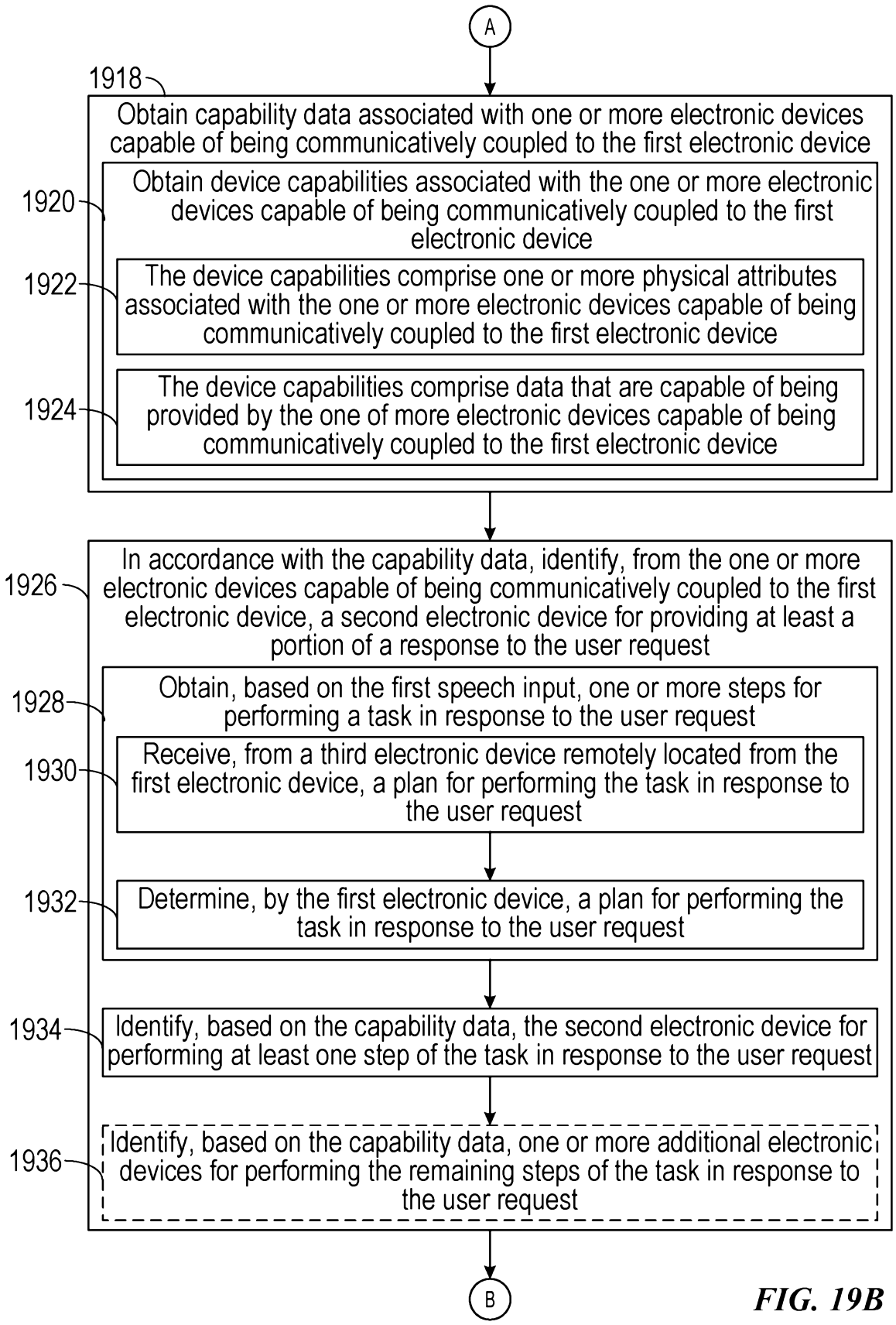

A
1918
Obtain capability data associated with one or more electronic devices capable of being communicatively coupled to the first electronic device
1920
Obtain device capabilities associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device
1922
The device capabilities comprise one or more physical attributes associated with the one or more electronic devices capable of being communicatively coupled to the first electronic device
1924
The device capabilities comprise data that are capable of being provided by the one of more electronic devices capable of being communicatively coupled to the first electronic device
1926
In accordance with the capability data, identify, from the one or more electronic devices capable of being communicatively coupled to the first electronic device, a second electronic device for providing at least a portion of a response to the user request
1928
Obtain, based on the first speech input, one or more steps for performing a task in response to the user request
1930
Receive, from a third electronic device remotely located from the first electronic device, a plan for performing the task in response to the user request
1932
Determine, by the first electronic device, a plan for performing the task in response to the user request
1934
Identify, based on the capability data, the second electronic device for performing at least one step of the task in response to the user request
1936
Identify, based on the capability data, one or more additional electronic devices for performing the remaining steps of the task in response to the user request
B

FIG. 19B

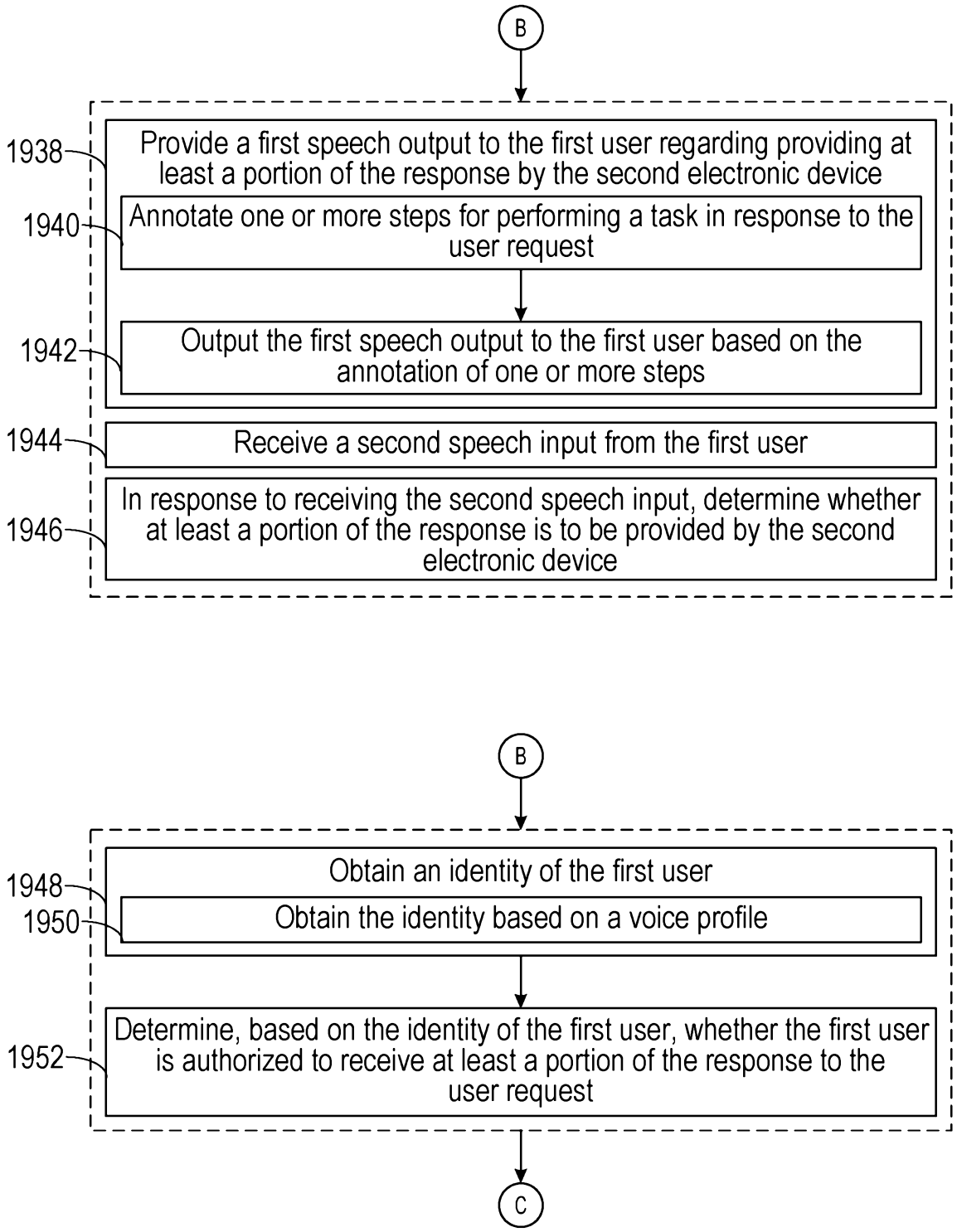

B
1938
Provide a first speech output to the first user regarding providing at least a portion of the response by the second electronic device
1940
Annotate one or more steps for performing a task in response to the user request
1942
Output the first speech output to the first user based on the annotation of one or more steps
1944
Receive a second speech input from the first user
1946
In response to receiving the second speech input, determine whether at least a portion of the response is to be provided by the second electronic device
B
1948
Obtain an identity of the first user
1950
Obtain the identity based on a voice profile
1952
Determine, based on the identity of the first user, whether the first user is authorized to receive at least a portion of the response to the user request
C

FIG. 19C

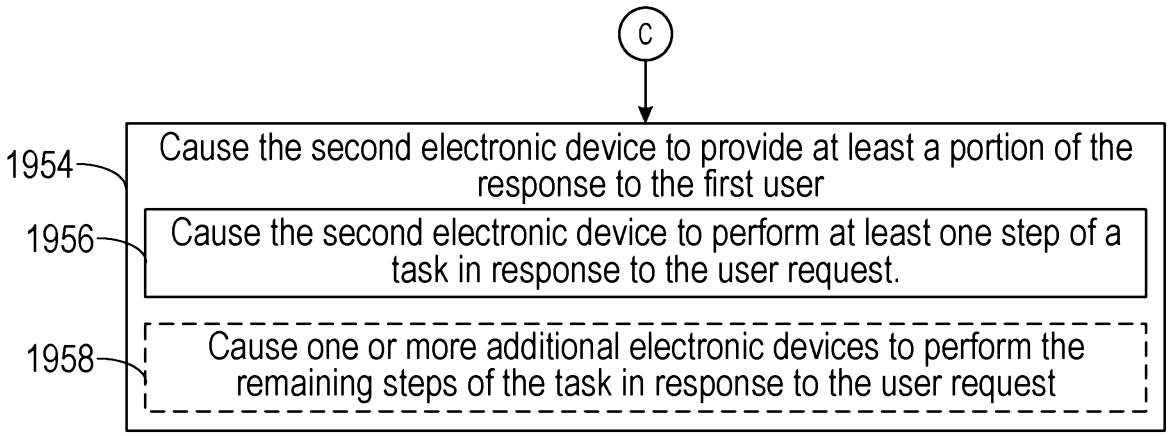

C
1954
Cause the second electronic device to provide at least a portion of the response to the first user
1956
Cause the second electronic device to perform at least one step of a task in response to the user request.
1958
Cause one or more additional electronic devices to perform the remaining steps of the task in response to the user request

FIG. 19D