US 20170358301A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2017/0358301 A1

(43) Pub. Date: Dec. 14, 2017

RAITIO et al.

(54) **DIGITAL ASSISTANT PROVIDING WHISPERED SPEECH**

(71) Applicant: **Apple Inc.**, Cupertino, CA (US)

(72) Inventors: **Tuomo J. RAITIO**, Sunnyvale, CA (US); **Melvyn J. HUNT**, Cheltenham (GB); **Hywel B. RICHARDS**, Rhiwbina (GB); **Madhusudan CHINTHAKUNTA**, Saratoga, CA (US)

(21) Appl. No.: **15/266,932**

(22) Filed: **Sep. 15, 2016**

Related U.S. Application Data

(60) Provisional application No. 62/348,705, filed on Jun. 10, 2016.

Publication Classification

(51) **Int. Cl.**

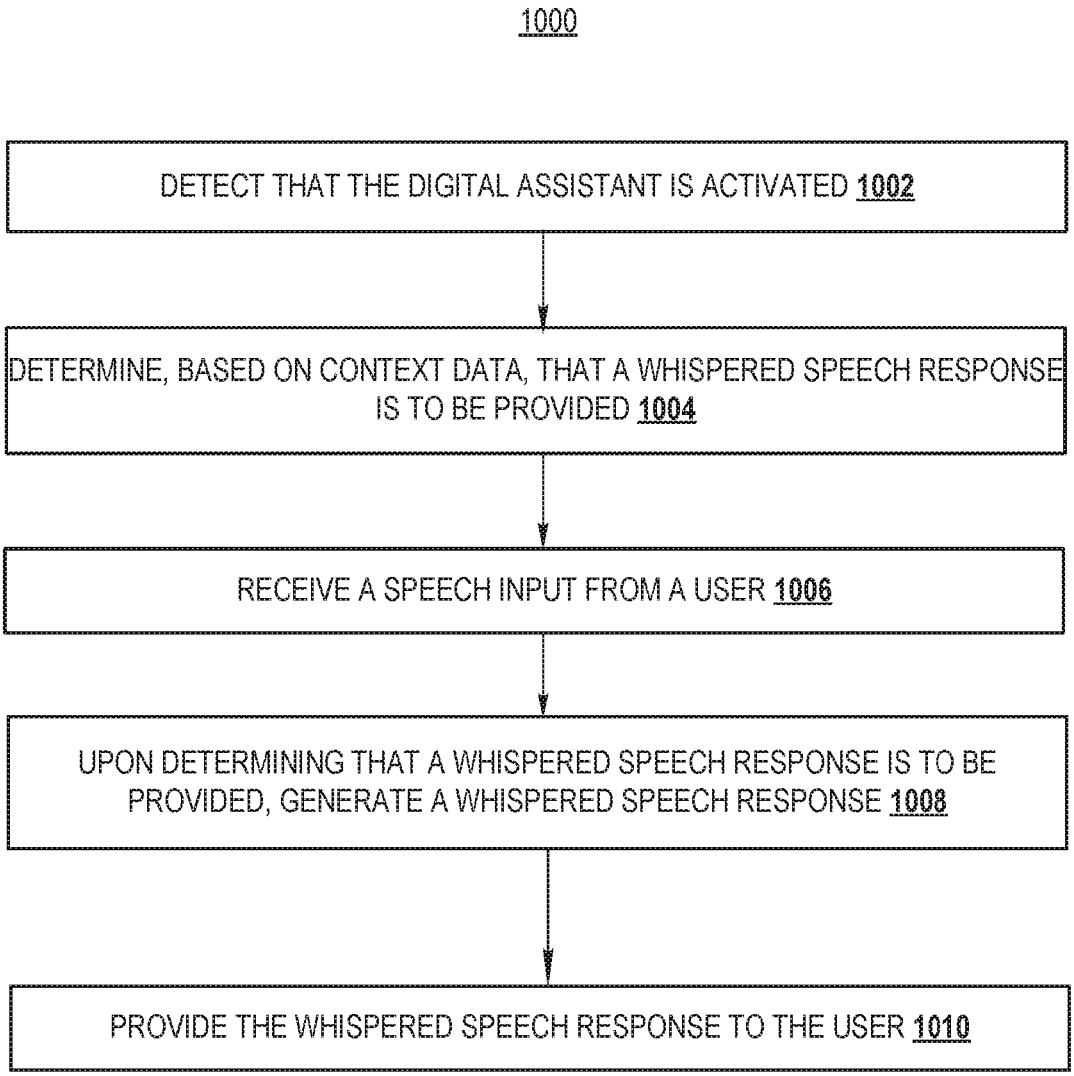
G10L 15/22	(2006.01)
G10L 25/24	(2013.01)
G10L 25/18	(2013.01)
G10L 13/033	(2013.01)

(52) **U.S. Cl.**

CPC ***G10L 15/22*** (2013.01); ***G10L 13/033*** (2013.01); ***G10L 25/24*** (2013.01); ***G10L 25/18*** (2013.01); *G10L 2015/223* (2013.01)

(57) **ABSTRACT**

Systems and processes for detecting and/or providing a whispered speech response are provided. In one example process, speech is received from a user, and based on the speech input, determined that a whispered speech response is to be provided. Upon determining that a whispered speech response is to be provided, the whispered speech response is generated and provided to the user.



Amazon v. SoundClear
US Patent 11,244,675
Amazon EX-1209

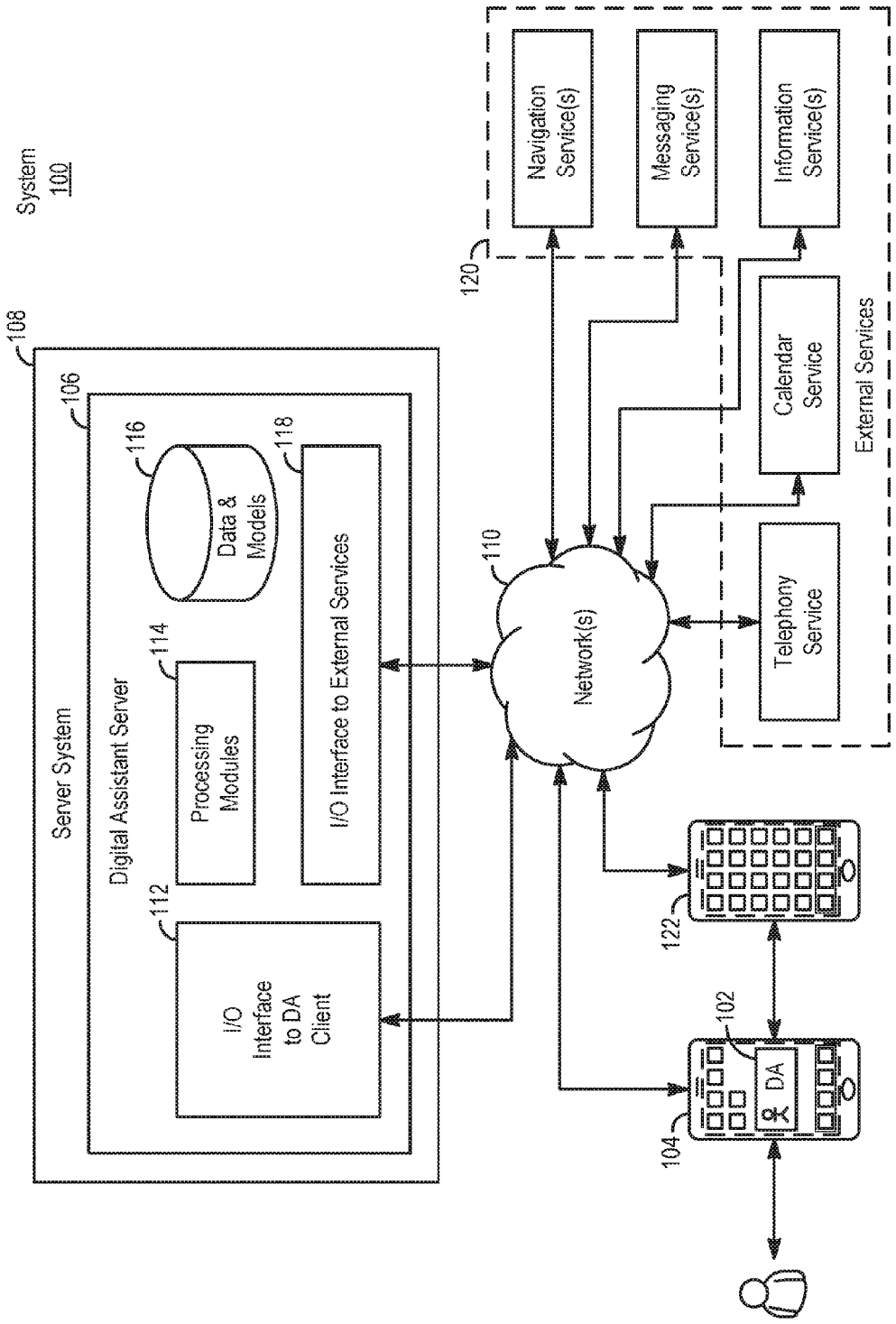

System
100
108
Server System
106
Digital Assistant Server
112
I/O Interface to DA Client
114
Processing Modules
116
Data & Models
118
I/O Interface to External Services
110
Network(s)
120
Navigation Service(s)
Messaging Service(s)
Information Service(s)
Calendar Service
Telephony Service
External Services
122
102
DA
104

FIG. 1

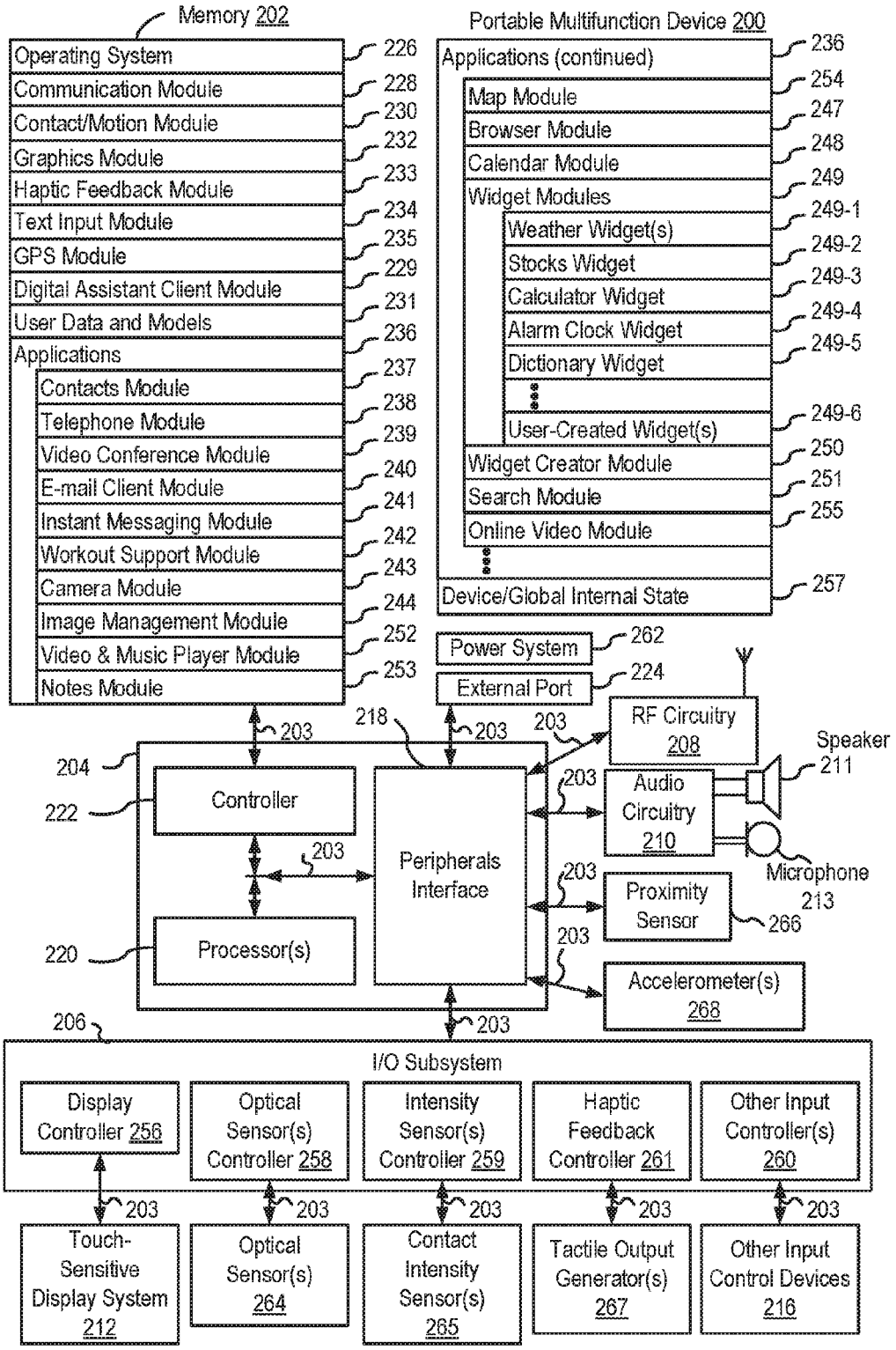

Memory 202
Operating System 226
Communication Module 228
Contact/Motion Module 230
Graphics Module 232
Haptic Feedback Module 233
Text Input Module 234
GPS Module 235
Digital Assistant Client Module 229
User Data and Models 231
Applications 236
Contacts Module 237
Telephone Module 238
Video Conference Module 239
E-mail Client Module 240
Instant Messaging Module 241
Workout Support Module 242
Camera Module 243
Image Management Module 244
Video & Music Player Module 252
Notes Module 253
Portable Multifunction Device 200
Applications (continued) 236
Map Module 254
Browser Module 247
Calendar Module 248
Widget Modules 249
Weather Widget(s) 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
Online Video Module 255
Device/Global Internal State 257
Power System 262
External Port 224
RF Circuitry 208
Speaker 211
Audio Circuitry 210
Microphone 213
Proximity Sensor 266
Accelerometer(s) 268
Controller 222
Processor(s) 220
Peripherals Interface 218
204
203
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

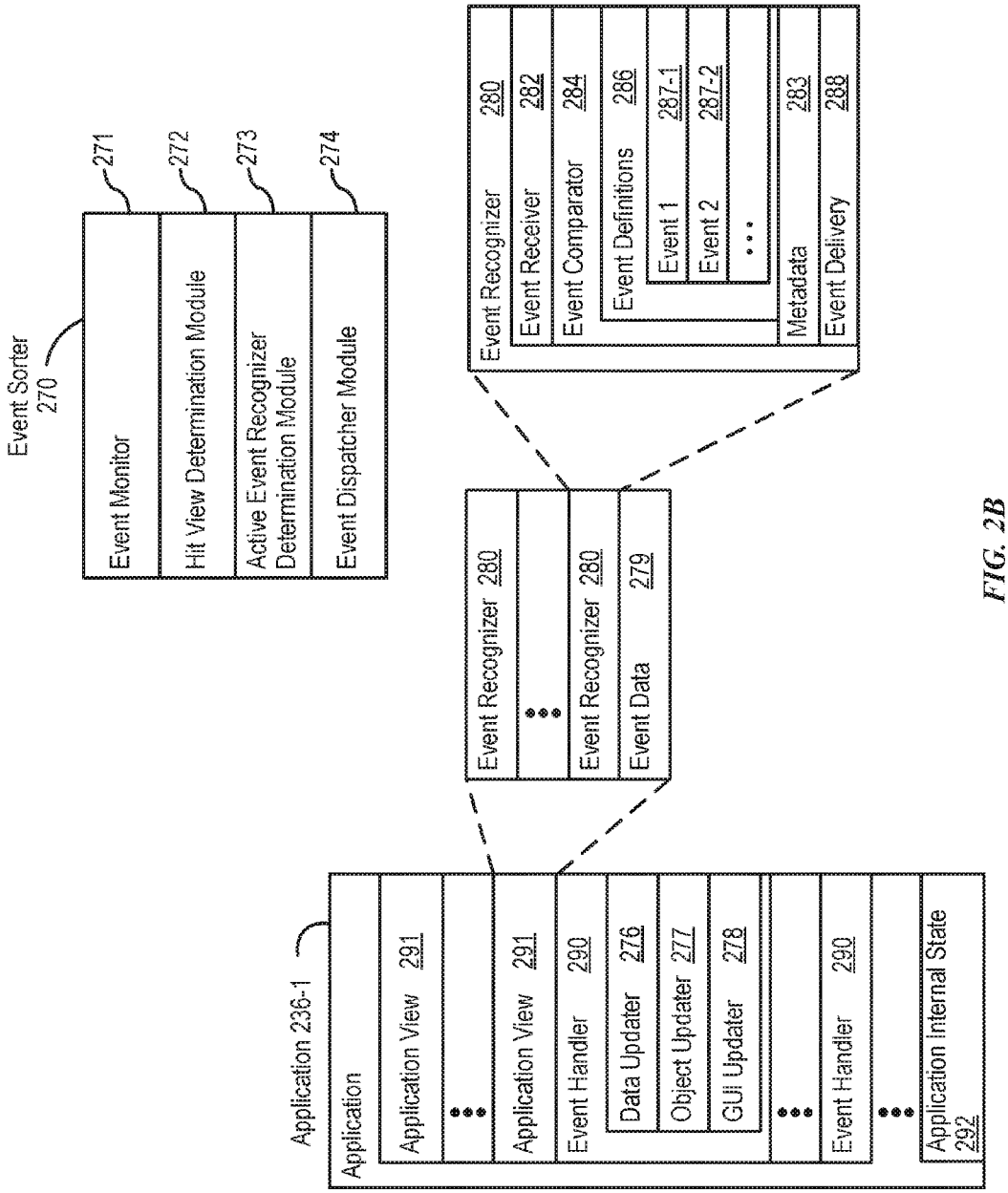

Event Sorter 270
271
Event Monitor
272
Hit View Determination Module
273
Active Event Recognizer Determination Module
274
Event Dispatcher Module
Application 236-1
Application
Application View 291
...
Application View 291
Event Handler 290
Data Updater 276
Object Updater 277
GUI Updater 278
...
Event Handler 290
...
Application Internal State 292
Event Recognizer 280
...
Event Recognizer 280
Event Data 279
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

FIG. 2B

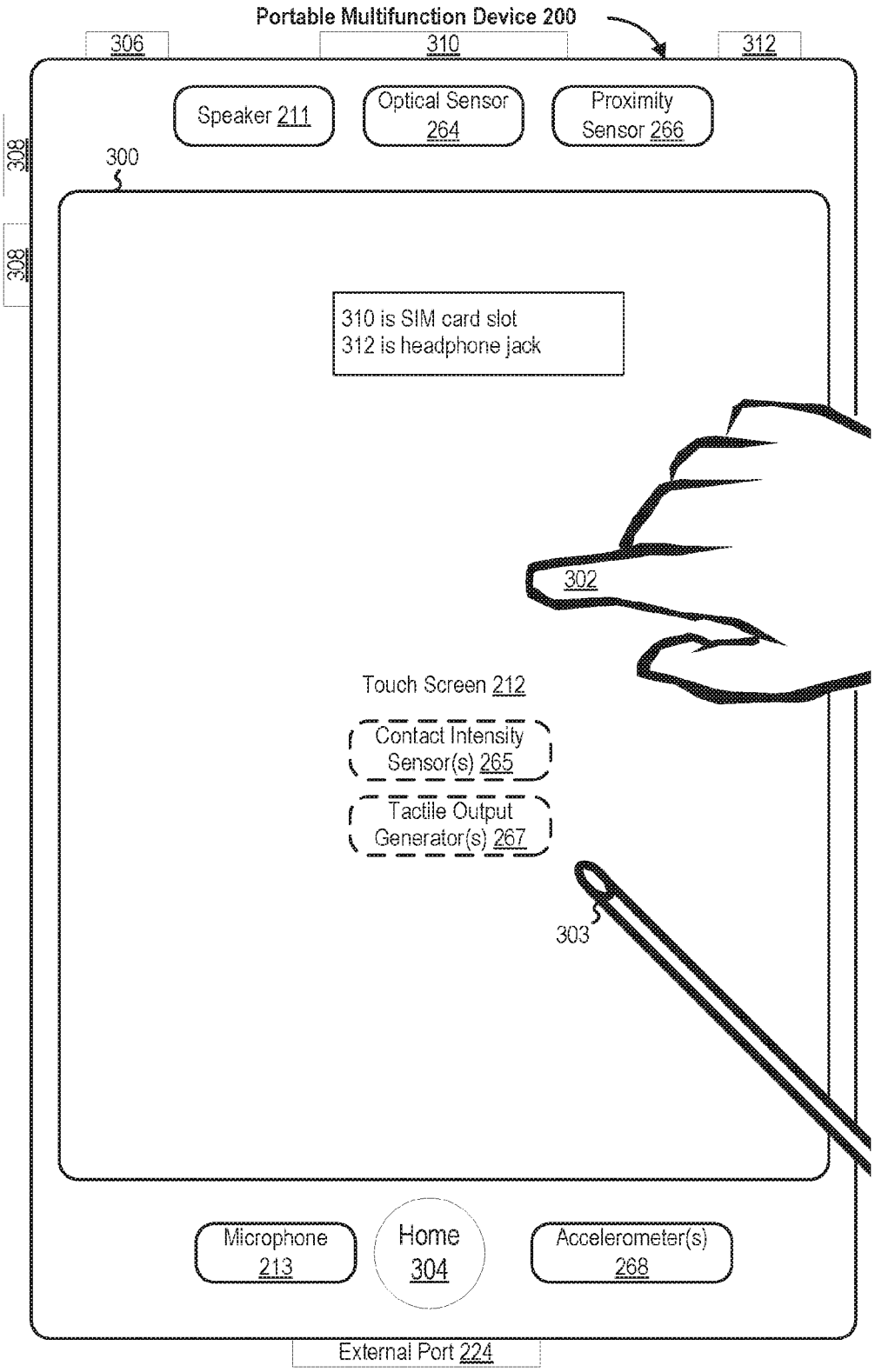

Portable Multifunction Device 200
306
310
312
Speaker 211
Optical Sensor 264
Proximity Sensor 266
308
300
308
310 is SIM card slot
312 is headphone jack
302
Touch Screen 212
Contact Intensity Sensor(s) 265
Tactile Output Generator(s) 267
303
Microphone 213
Home 304
Accelerometer(s) 268
External Port 224

FIG. 3

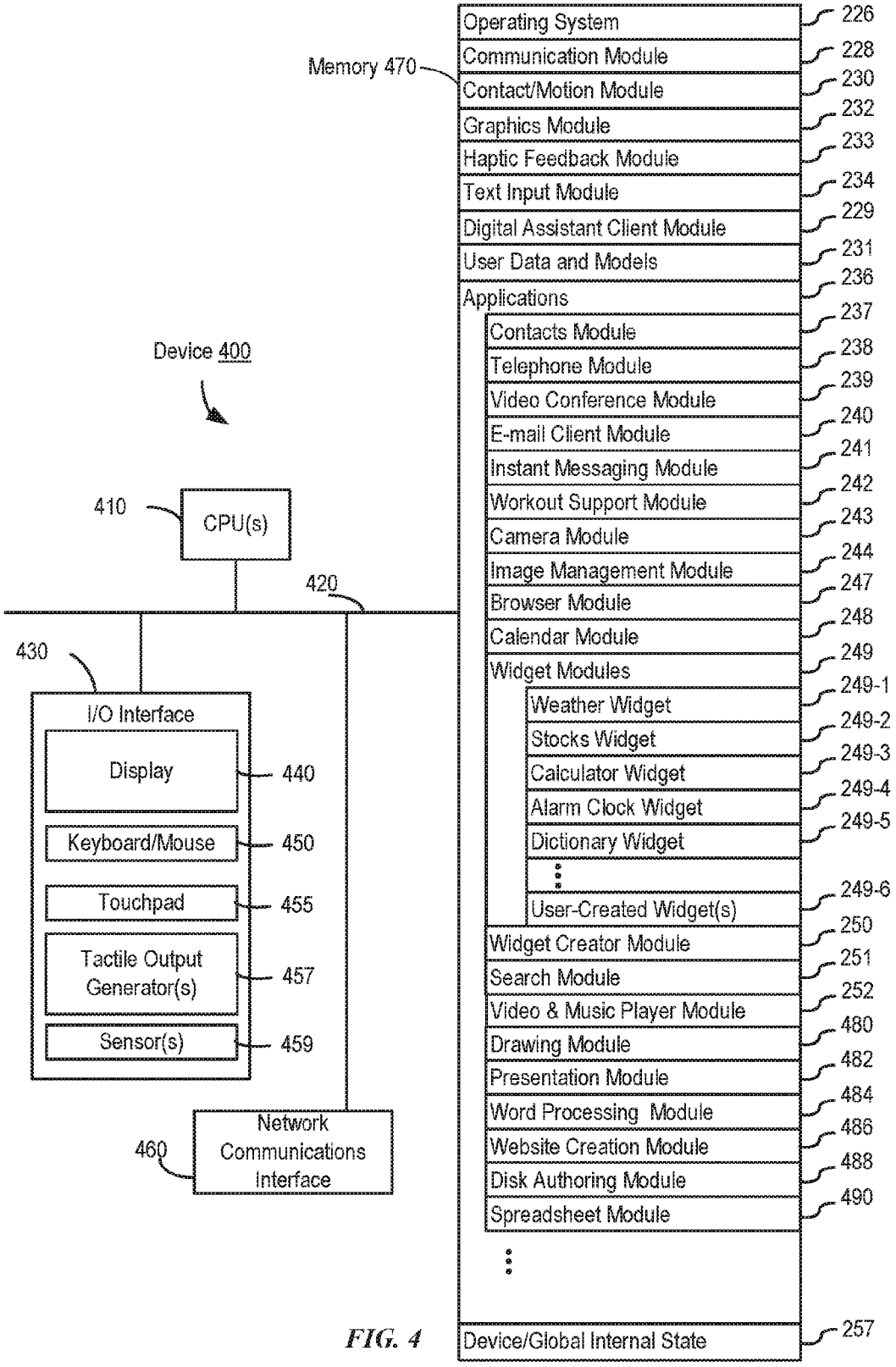

Memory 470
Operating System 226
Communication Module 228
Contact/Motion Module 230
Graphics Module 232
Haptic Feedback Module 233
Text Input Module 234
Digital Assistant Client Module 229
User Data and Models 231
Applications 236
Contacts Module 237
Telephone Module 238
Video Conference Module 239
E-mail 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
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

FIG. 4

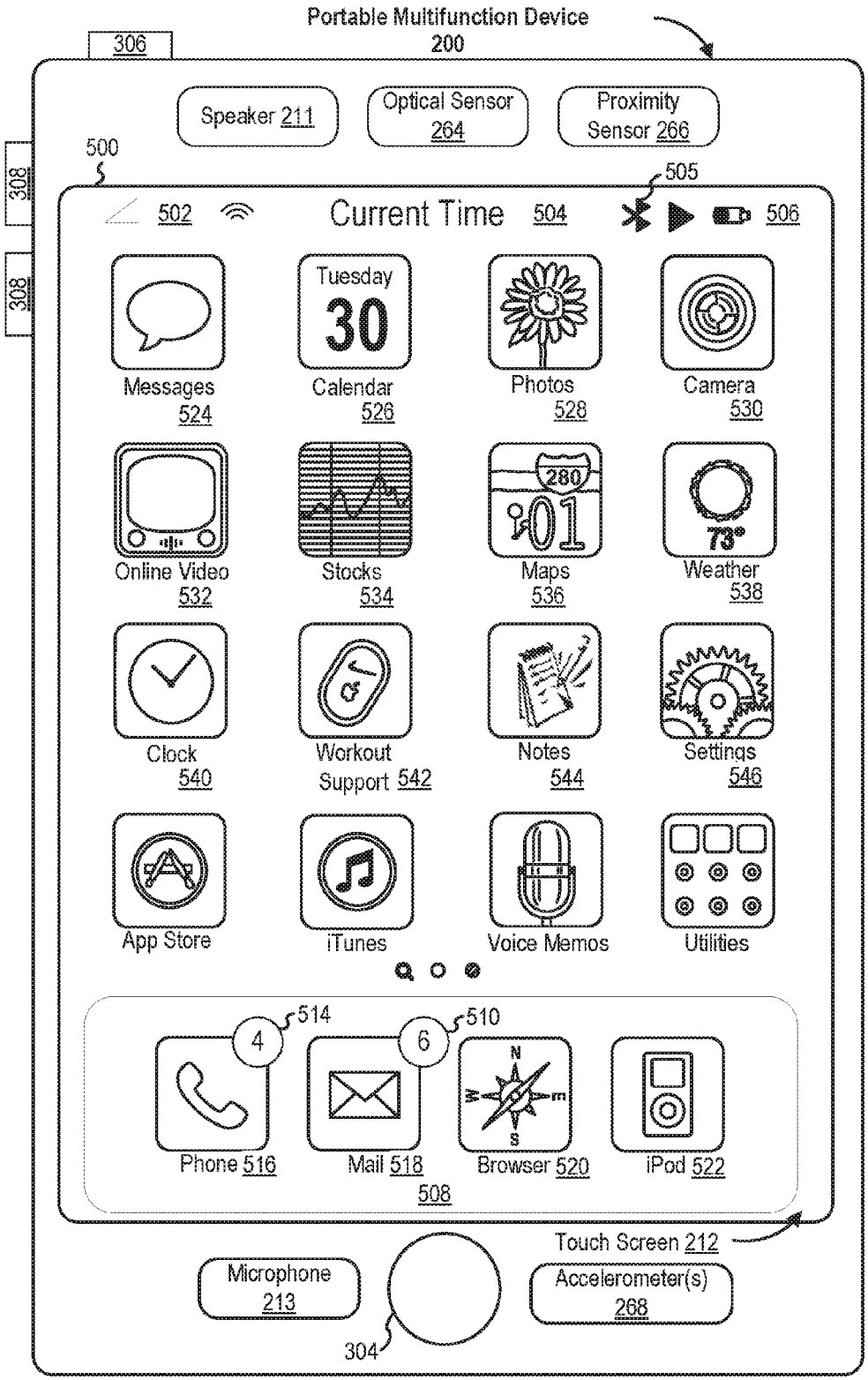
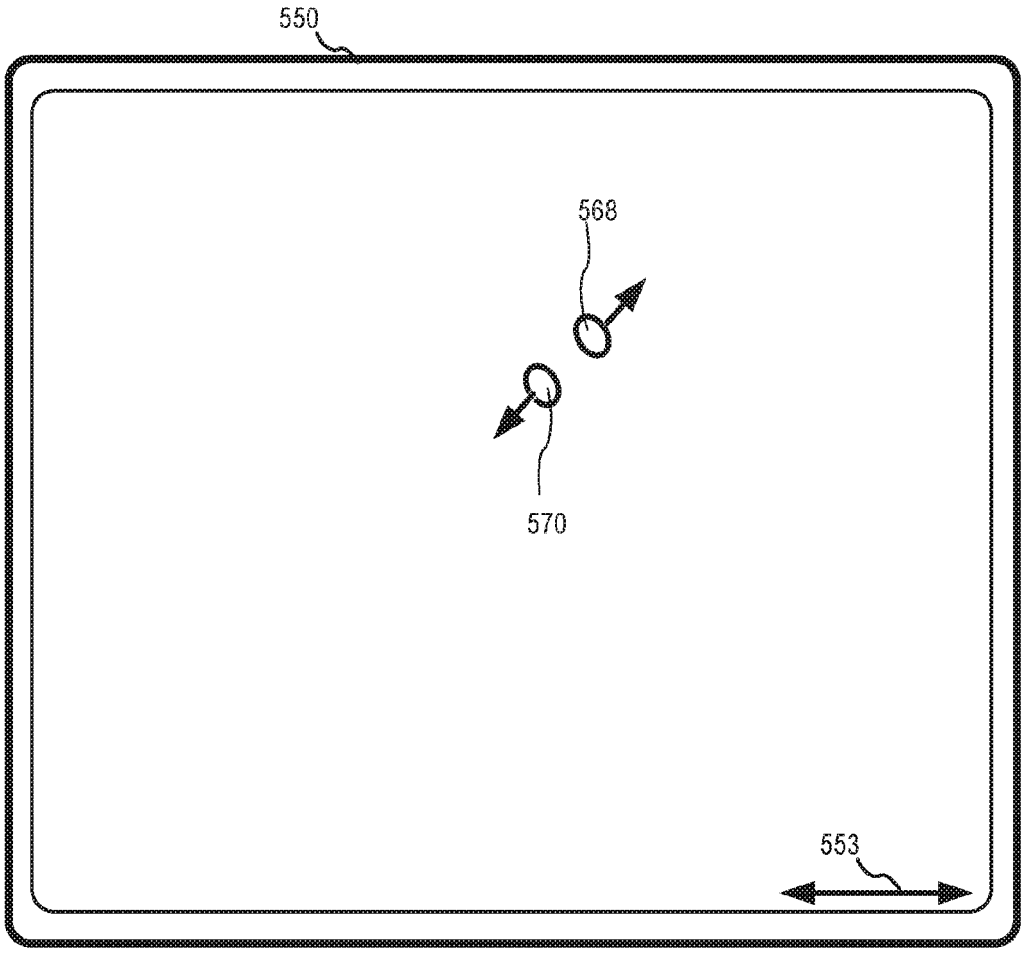

Portable Multifunction Device
200
306
Speaker 211
Optical Sensor 264
Proximity Sensor 266
500
308
308
505
502
Current Time 504
506
Tuesday
30
Messages 524
Calendar 526
Photos 528
Camera 530
Online Video 532
Stocks 534
Maps 536
Weather 538
Clock 540
Workout Support 542
Notes 544
Settings 546
App Store
iTunes
Voice Memos
Utilities
514
510
4
6
Phone 516
Mail 518
Browser 520
iPod 522
508
Touch Screen 212
Microphone 213
Accelerometer(s) 268
304

FIG. 5A


550
568
570
553

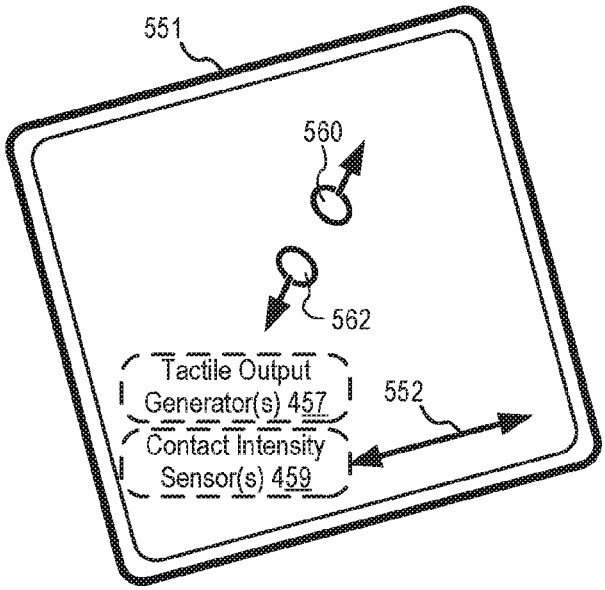

551
560
562
Tactile Output Generator(s) 457
552
Contact Intensity Sensor(s) 459

FIG. 5B

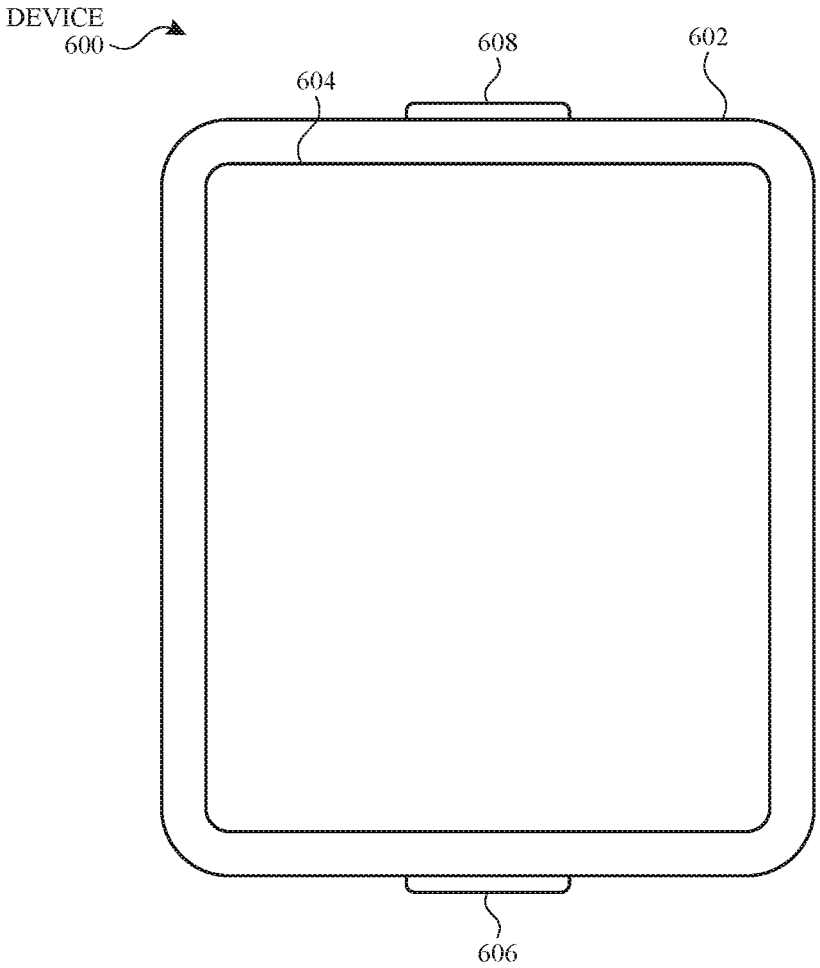


FIG. 6A

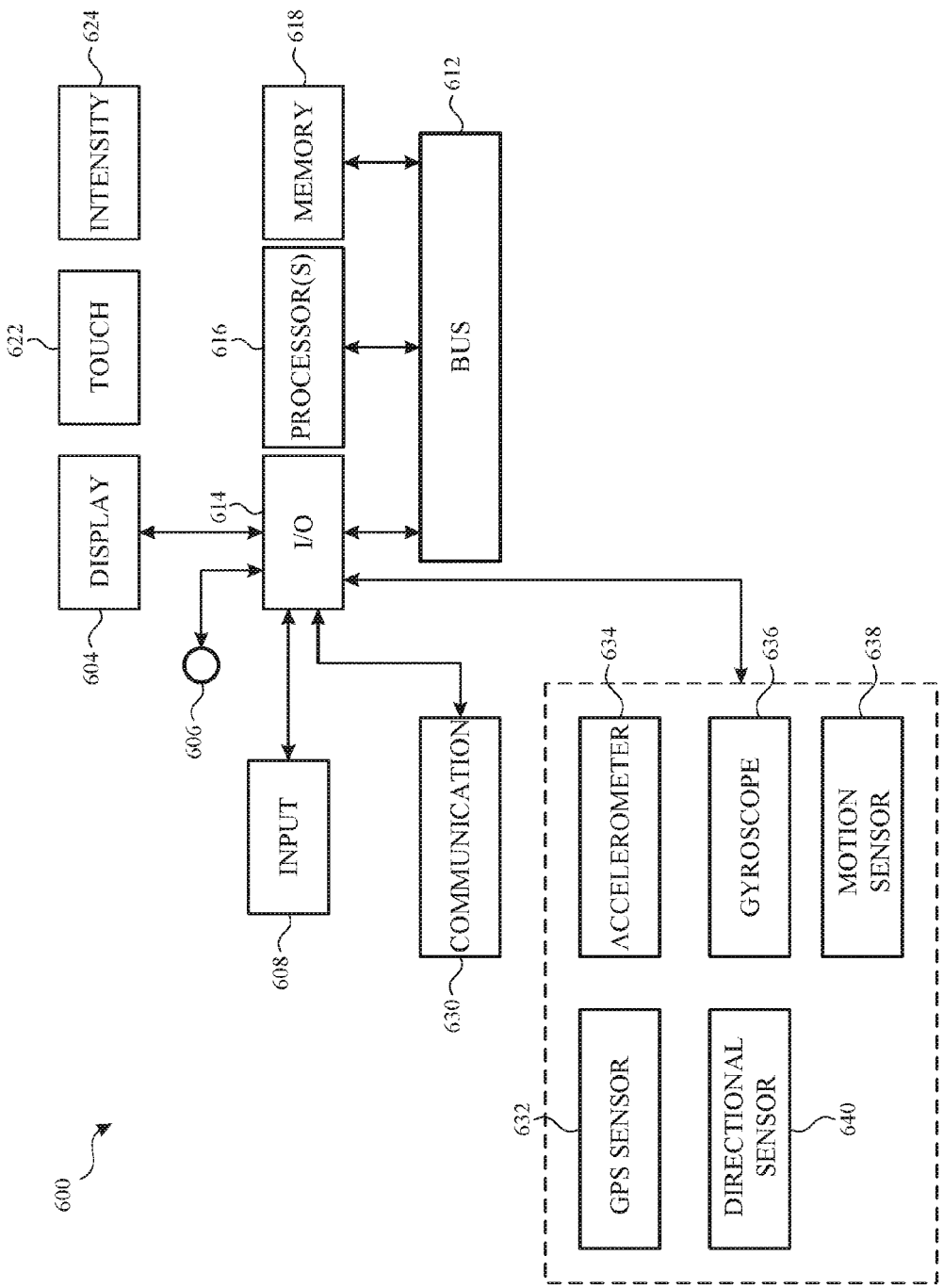

600
DISPLAY
604
606
INPUT
608
TOUCH
622
INTENSITY
624
I/O
614
PROCESSOR(S)
616
MEMORY
618
BUS
612
COMMUNICATION
630
GPS SENSOR
632
ACCELEROMETER
634
DIRECTIONAL SENSOR
640
GYROSCOPE
636
MOTION SENSOR
638

FIG. 6B

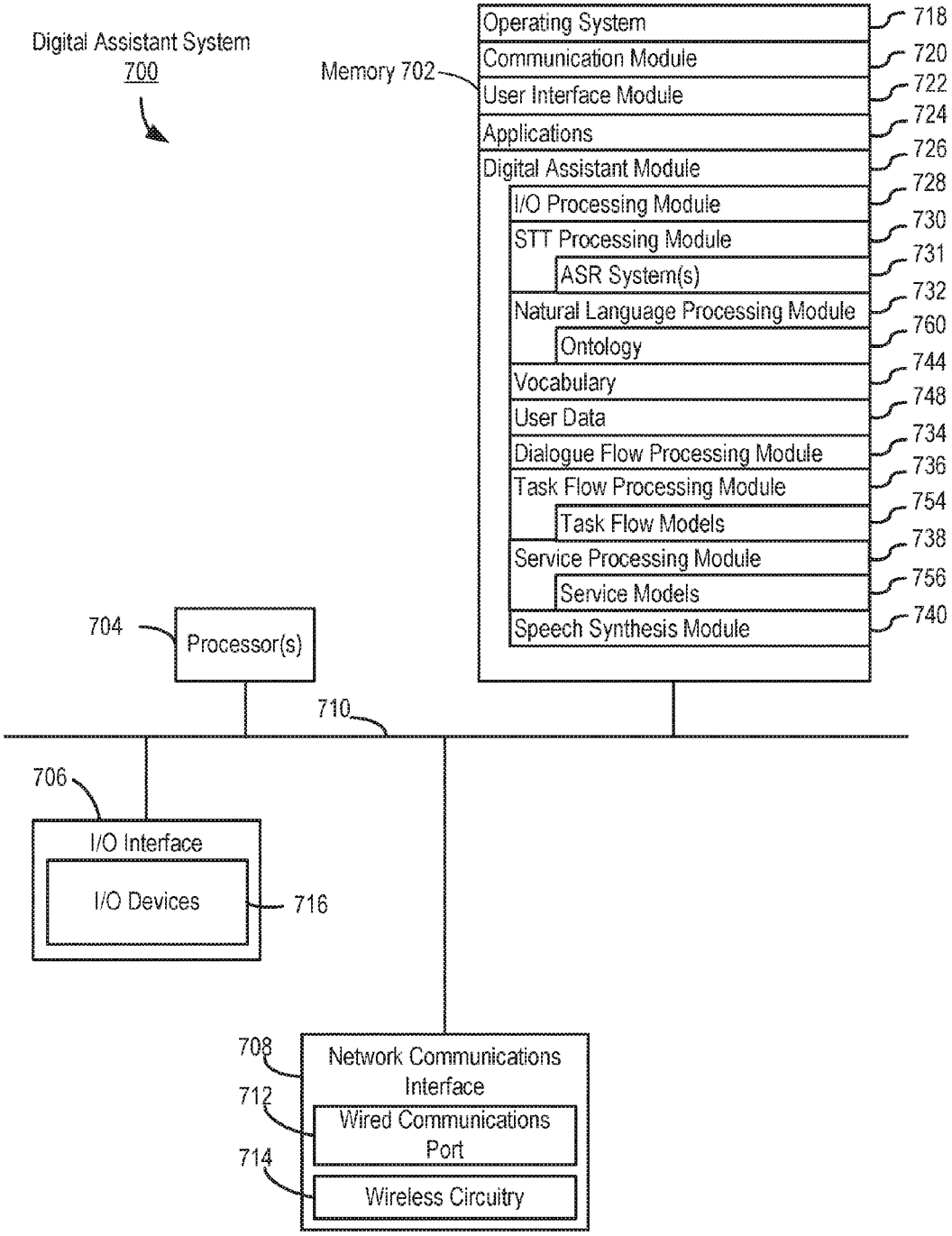

Digital Assistant System
700
Memory 702
Operating System
Communication Module
User Interface Module
Applications
Digital Assistant Module
I/O Processing Module
STT Processing Module
ASR System(s)
Natural Language Processing Module
Ontology
Vocabulary
User Data
Dialogue Flow Processing Module
Task Flow Processing Module
Task Flow Models
Service Processing Module
Service Models
Speech Synthesis Module
718
720
722
724
726
728
730
731
732
760
744
748
734
736
754
738
756
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

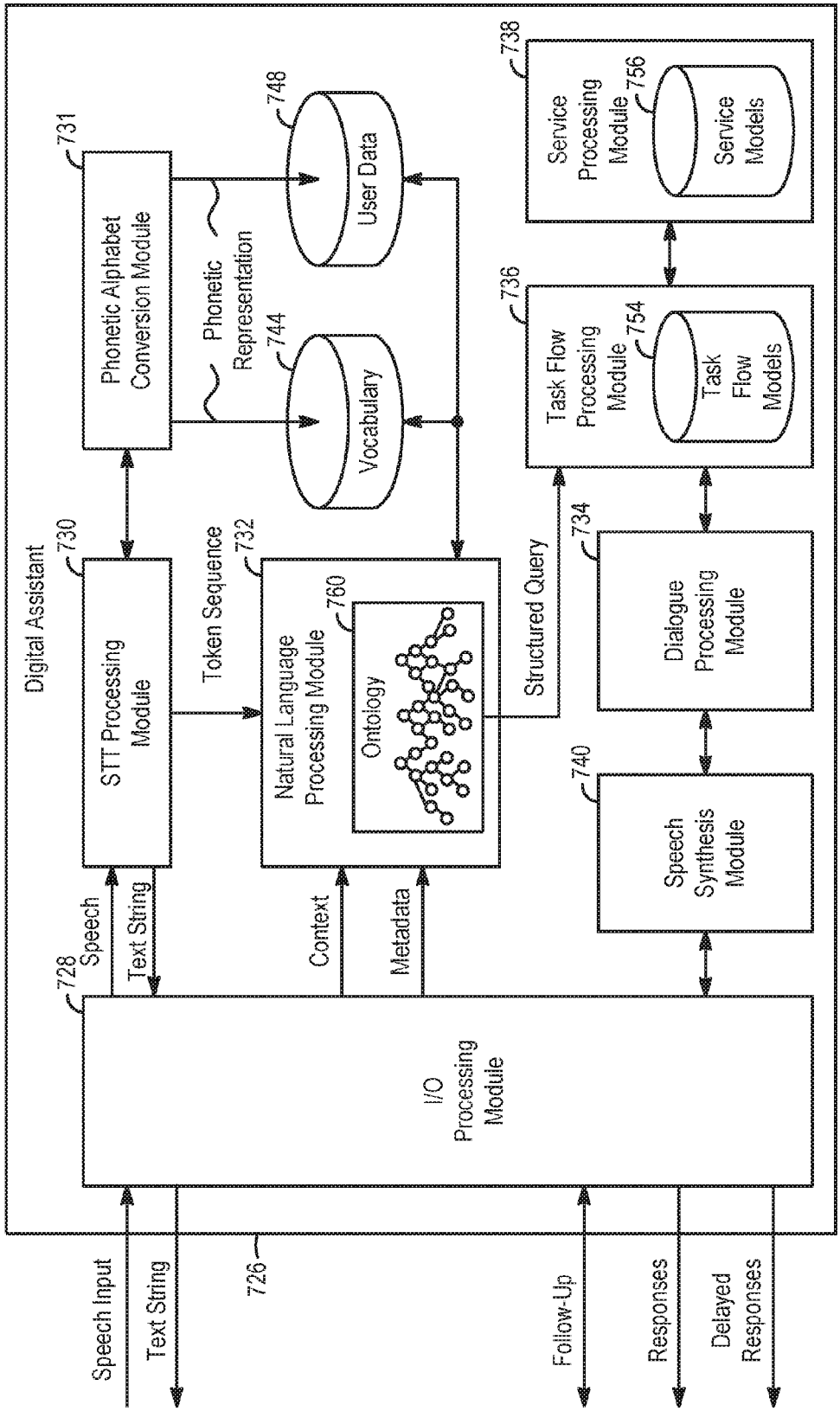

Digital Assistant
730
STT Processing Module
731
Phonetic Alphabet Conversion Module
Phonetic Representation
748
User Data
744
Vocabulary
Token Sequence
732
Natural Language Processing Module
760
Ontology
Structured Query
736
Task Flow Processing Module
754
Task Flow Models
738
Service Processing Module
756
Service Models
734
Dialogue Processing Module
740
Speech Synthesis Module
728
I/O Processing Module
Speech
Text String
Context
Metadata
Speech Input
Text String
726
Follow-Up
Responses
Delayed Responses

FIG. 7B

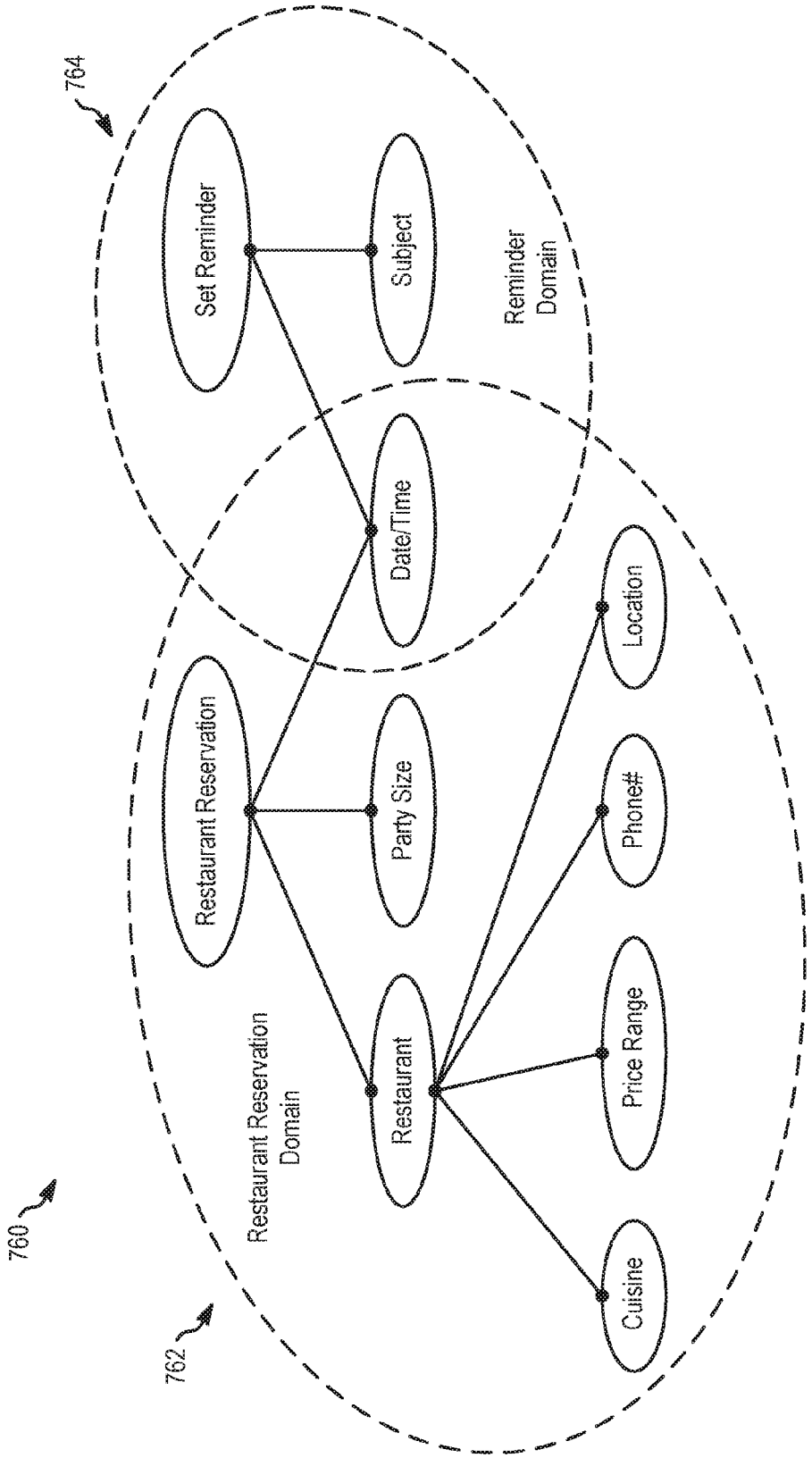

760
762
764
Restaurant Reservation Domain
Reminder Domain
Restaurant Reservation
Set Reminder
Restaurant
Date/Time
Party Size
Subject
Cuisine
Price Range
Phone#
Location

FIG. 7C

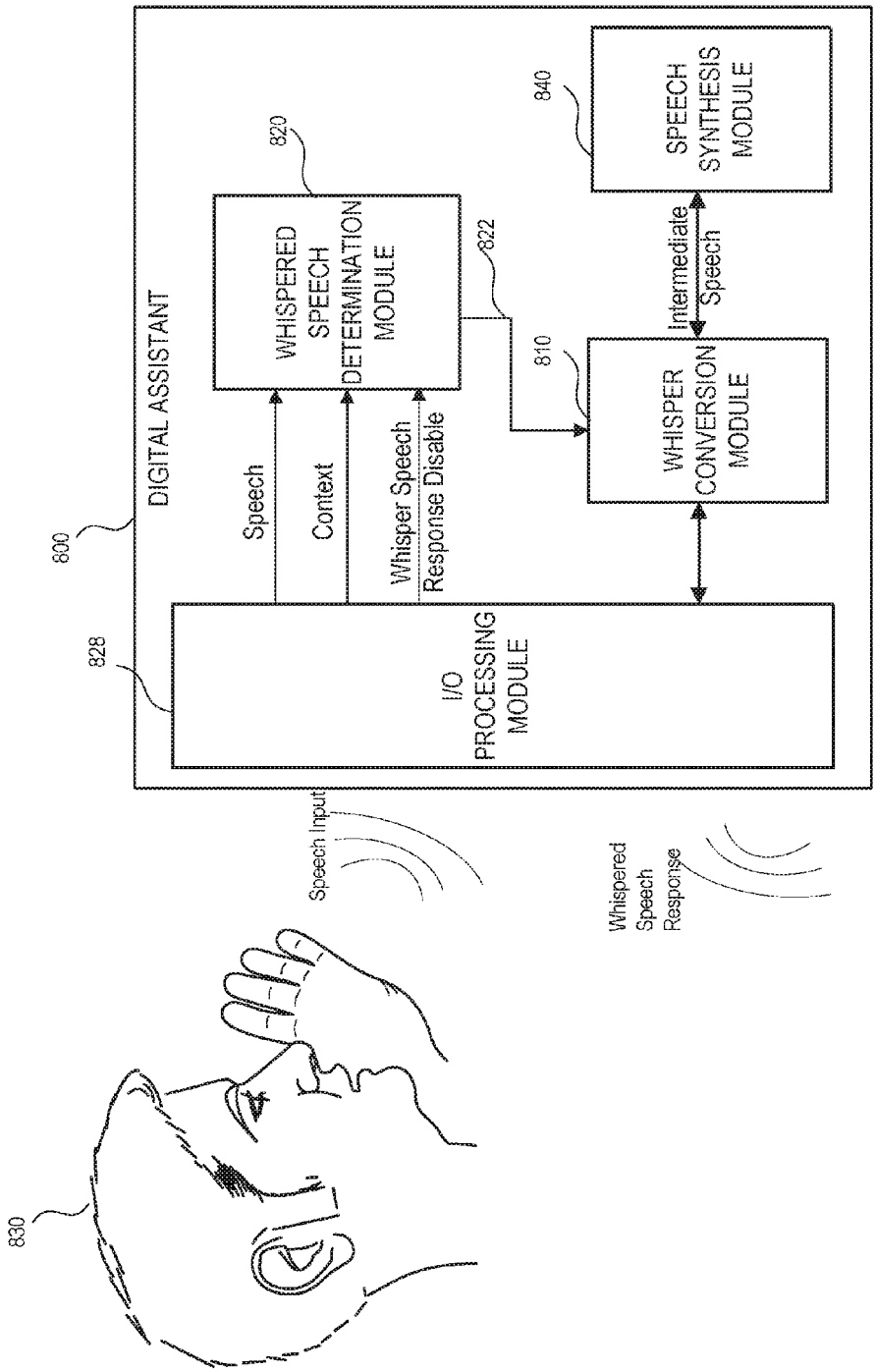

800
DIGITAL ASSISTANT
828
I/O PROCESSING MODULE
Speech
Context
Whisper Speech Response Disable
820
WHISPERED SPEECH DETERMINATION MODULE
822
810
WHISPER CONVERSION MODULE
Intermediate Speech
840
SPEECH SYNTHESIS MODULE
Speech Input
Whispered Speech Response
830

FIG. 8A

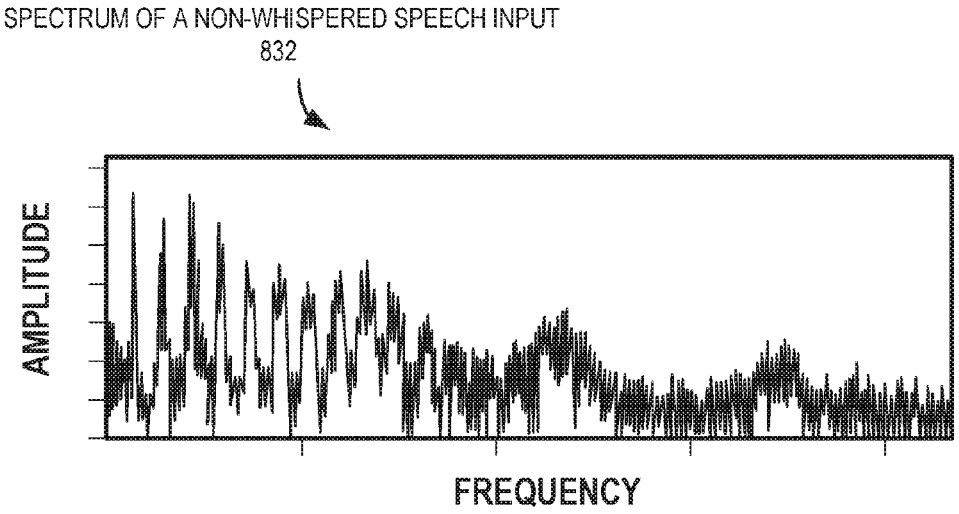

SPECTRUM OF A NON-WHISPERED SPEECH INPUT
832
AMPLITUDE
FREQUENCY

FIG. 8B

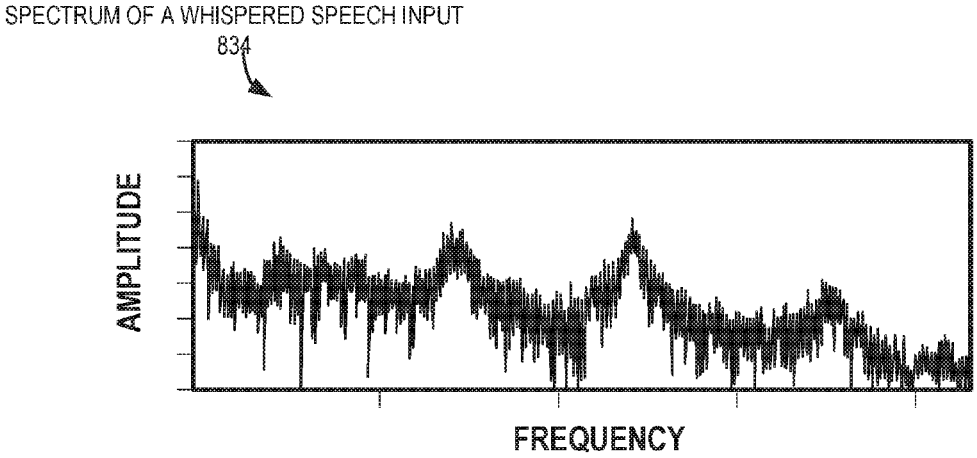

SPECTRUM OF A WHISPERED SPEECH INPUT
834
AMPLITUDE
FREQUENCY

FIG. 8C

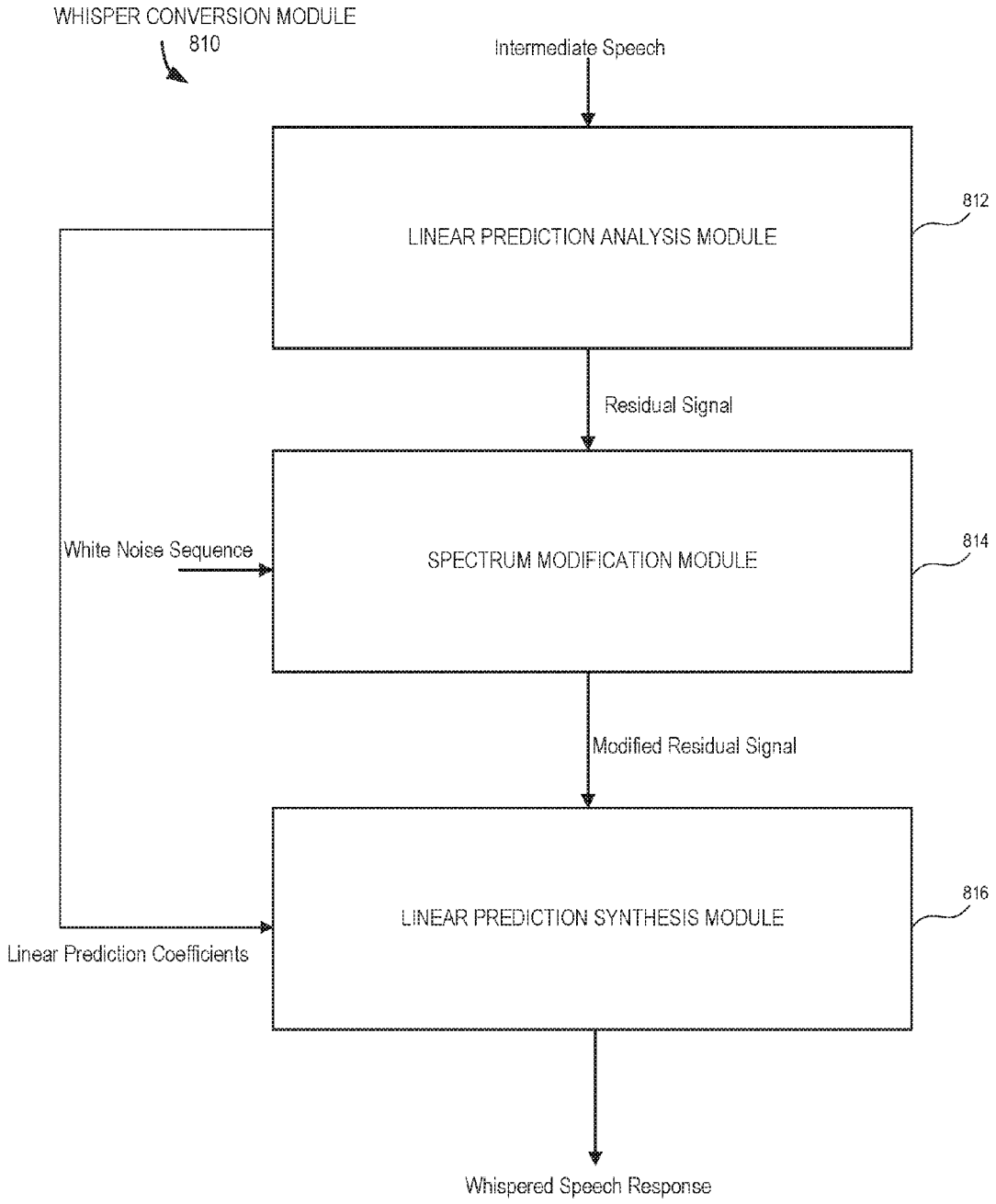

WHISPER CONVERSION MODULE
810
Intermediate Speech
LINEAR PREDICTION ANALYSIS MODULE
812
Residual Signal
White Noise Sequence
SPECTRUM MODIFICATION MODULE
814
Modified Residual Signal
LINEAR PREDICTION SYNTHESIS MODULE
816
Linear Prediction Coefficients
Whispered Speech Response

FIG. 8D

900
RECEIVE A SPEECH INPUT FROM A USER 902
THE SPEECH INPUT COMPRISES AN INFORMATIONAL REQUEST 904
THE SPEECH INPUT COMPRISES A REQUEST TO PERFORM A TASK 906
DETERMINE WHETHER PROVIDING A WHISPERED SPEECH RESPONSE IS DISABLED 908
IN ACCORDANCE WITH THE DETERMINATION THAT PROVIDING THE WHISPERED SPEECH RESPONSE IS DISABLED, GENERATE A NON-WHISPERED SPEECH RESPONSE, AND PROVIDE THE NON-WHISPERED SPEECH RESPONSE TO THE USER IN LIEU OF THE WHISPERED SPEECH RESPONSE 910
DETERMINE, BASED ON THE SPEECH INPUT, THAT A WHISPERED SPEECH RESPONSE IS TO BE PROVIDED 912
THE WHISPERED SPEECH RESPONSE COMPRISES A RESPONSE TO THE INFORMATIONAL REQUEST 913
THE WHISPERED SPEECH RESPONSE COMPRISES A RESPONSE ASSOCIATED WITH PERFORMING THE TASK 914
DETERMINE WHETHER THE SPEECH INPUT INCLUDES A WHISPERED SPEECH INPUT 916
THE WHISPERED SPEECH INPUT IS ASSOCIATED WITH A FIRST SPECTRUM HAVING ONE OR MORE FIRST SPECTRUM CHARACTERISTICS ASSOCIATED WITH A WHISPERED SPEECH 918
THE ONE OR MORE FIRST SPECTRUM CHARACTERISTICS COMPRISES A FIRST AMPLITUDE 919
THE ONE OR MORE FIRST SPECTRUM CHARACTERISTICS COMPRISES A FIRST ENERGY 920
THE ONE OR MORE FIRST SPECTRUM CHARACTERISTICS COMPRISES A FIRST VOLUME 921
THE ONE OR MORE FIRST SPECTRUM CHARACTERISTICS COMPRISES A FIRST SLOPE OF THE FIRST SPECTRUM 922

FIG. 9A

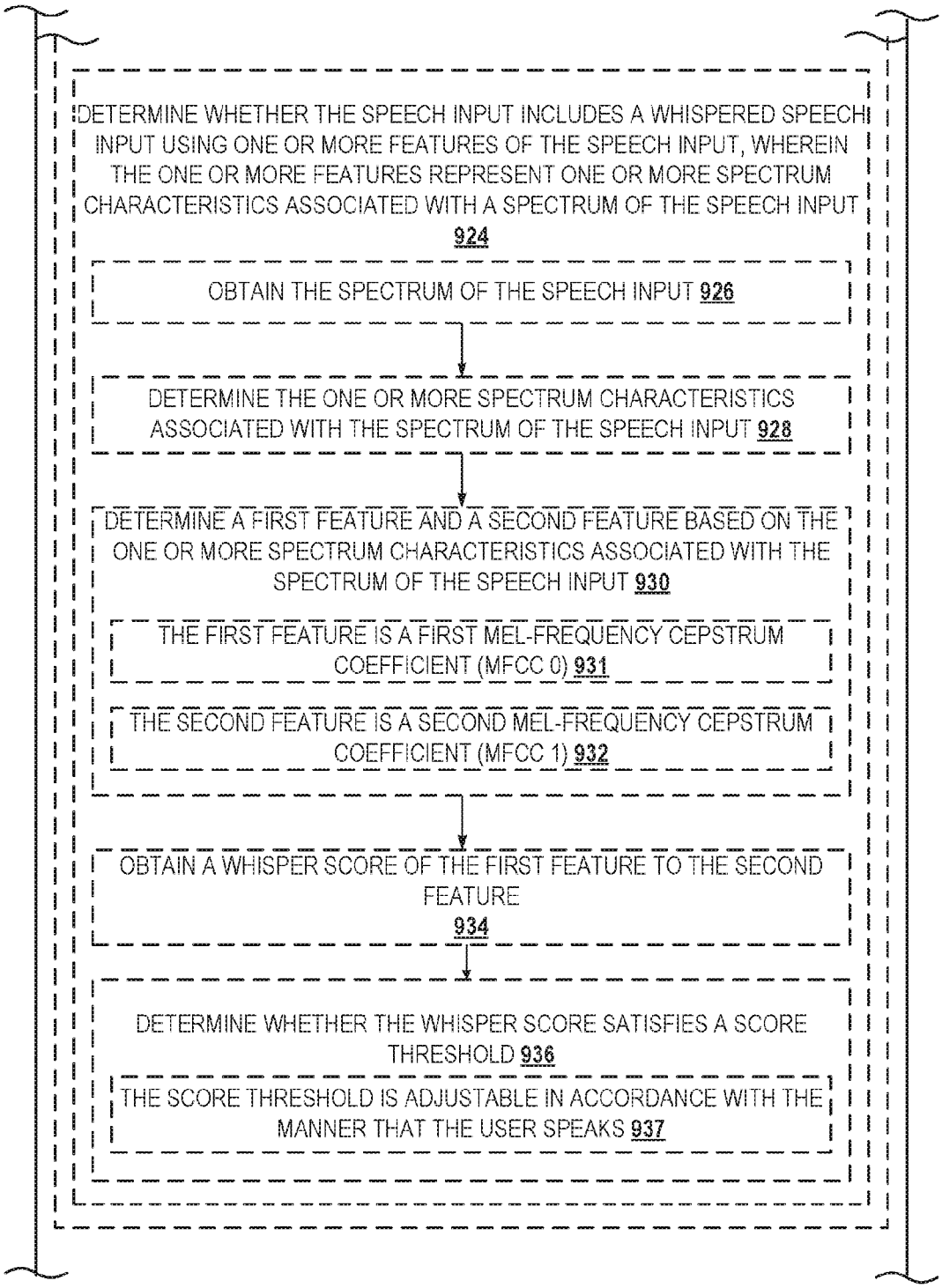

DETERMINE WHETHER THE SPEECH INPUT INCLUDES A WHISPERED SPEECH INPUT USING ONE OR MORE FEATURES OF THE SPEECH INPUT, WHEREIN THE ONE OR MORE FEATURES REPRESENT ONE OR MORE SPECTRUM CHARACTERISTICS ASSOCIATED WITH A SPECTRUM OF THE SPEECH INPUT 924
OBTAIN THE SPECTRUM OF THE SPEECH INPUT 926
DETERMINE THE ONE OR MORE SPECTRUM CHARACTERISTICS ASSOCIATED WITH THE SPECTRUM OF THE SPEECH INPUT 928
DETERMINE A FIRST FEATURE AND A SECOND FEATURE BASED ON THE ONE OR MORE SPECTRUM CHARACTERISTICS ASSOCIATED WITH THE SPECTRUM OF THE SPEECH INPUT 930
THE FIRST FEATURE IS A FIRST MEL-FREQUENCY CEPSTRUM COEFFICIENT (MFCC 0) 931
THE SECOND FEATURE IS A SECOND MEL-FREQUENCY CEPSTRUM COEFFICIENT (MFCC 1) 932
OBTAIN A WHISPER SCORE OF THE FIRST FEATURE TO THE SECOND FEATURE 934
DETERMINE WHETHER THE WHISPER SCORE SATISFIES A SCORE THRESHOLD 936
THE SCORE THRESHOLD IS ADJUSTABLE IN ACCORDANCE WITH THE MANNER THAT THE USER SPEAKS 937

FIG. 9B

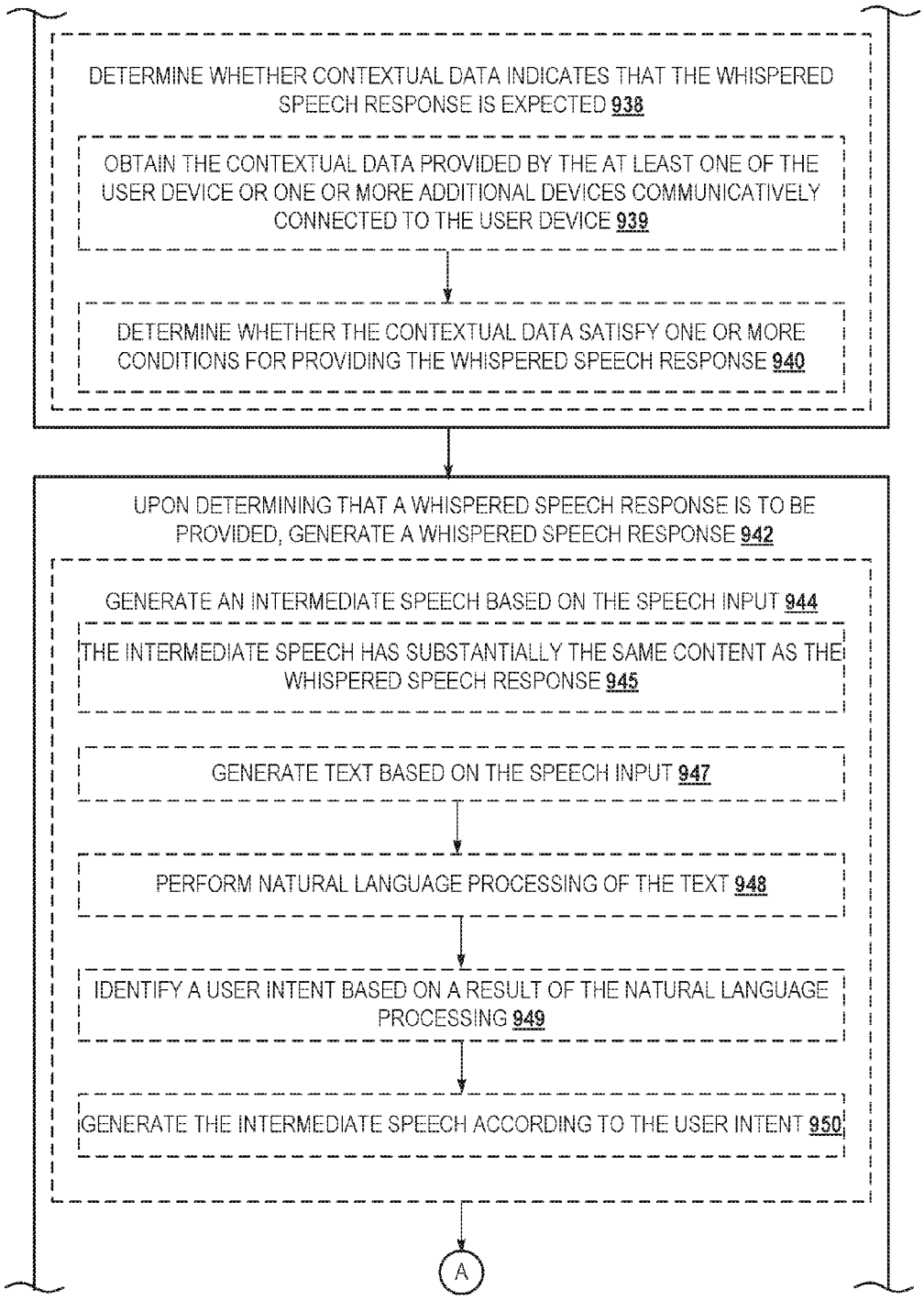

DETERMINE WHETHER CONTEXTUAL DATA INDICATES THAT THE WHISPERED SPEECH RESPONSE IS EXPECTED 938
OBTAIN THE CONTEXTUAL DATA PROVIDED BY THE AT LEAST ONE OF THE USER DEVICE OR ONE OR MORE ADDITIONAL DEVICES COMMUNICATIVELY CONNECTED TO THE USER DEVICE 939
DETERMINE WHETHER THE CONTEXTUAL DATA SATISFY ONE OR MORE CONDITIONS FOR PROVIDING THE WHISPERED SPEECH RESPONSE 940
UPON DETERMINING THAT A WHISPERED SPEECH RESPONSE IS TO BE PROVIDED, GENERATE A WHISPERED SPEECH RESPONSE 942
GENERATE AN INTERMEDIATE SPEECH BASED ON THE SPEECH INPUT 944
THE INTERMEDIATE SPEECH HAS SUBSTANTIALLY THE SAME CONTENT AS THE WHISPERED SPEECH RESPONSE 945
GENERATE TEXT BASED ON THE SPEECH INPUT 947
PERFORM NATURAL LANGUAGE PROCESSING OF THE TEXT 948
IDENTIFY A USER INTENT BASED ON A RESULT OF THE NATURAL LANGUAGE PROCESSING 949
GENERATE THE INTERMEDIATE SPEECH ACCORDING TO THE USER INTENT 950
A

FIG. 9C

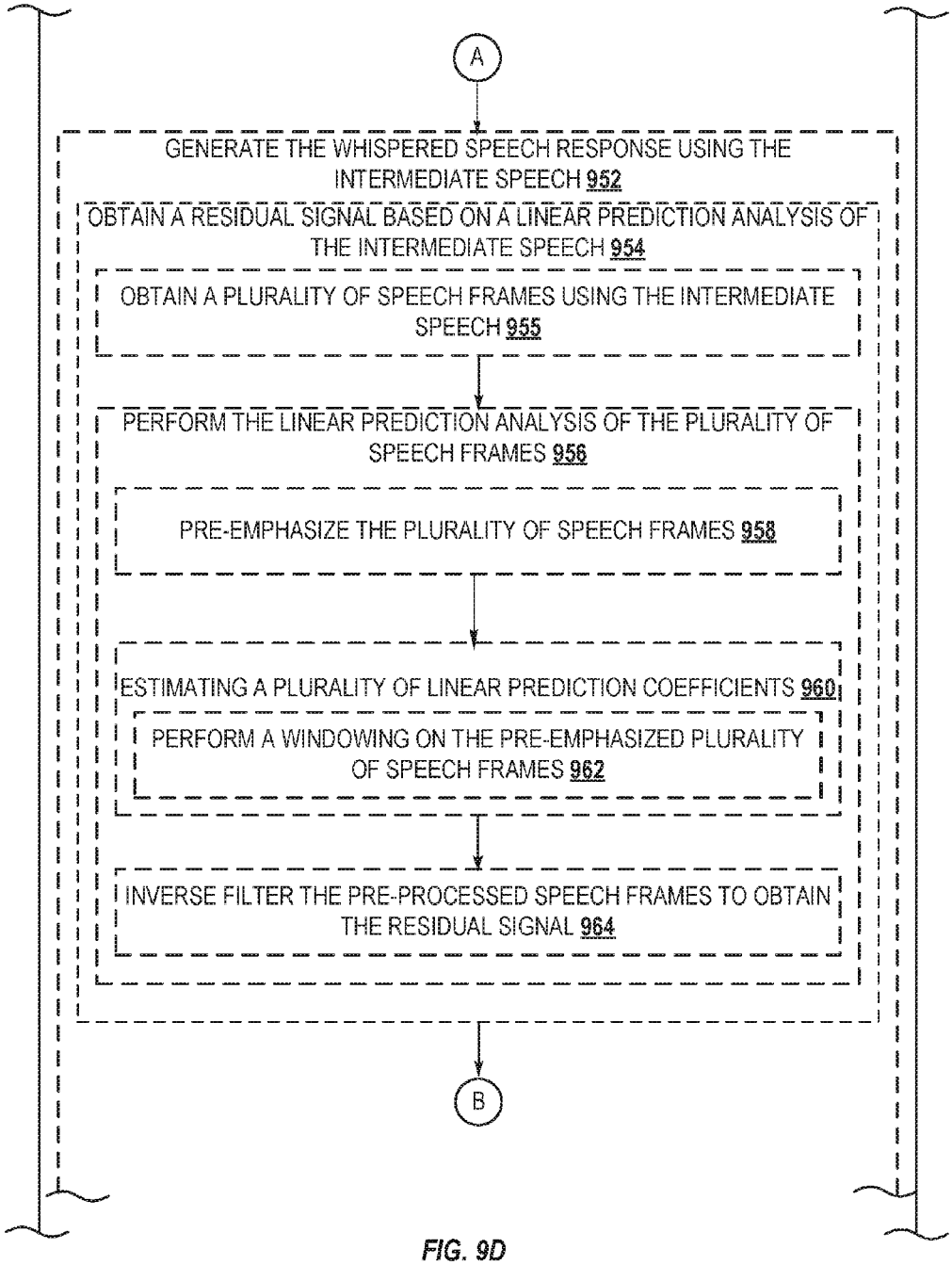

A
GENERATE THE WHISPERED SPEECH RESPONSE USING THE INTERMEDIATE SPEECH 952
OBTAIN A RESIDUAL SIGNAL BASED ON A LINEAR PREDICTION ANALYSIS OF THE INTERMEDIATE SPEECH 954
OBTAIN A PLURALITY OF SPEECH FRAMES USING THE INTERMEDIATE SPEECH 955
PERFORM THE LINEAR PREDICTION ANALYSIS OF THE PLURALITY OF SPEECH FRAMES 956
PRE-EMPHASIZE THE PLURALITY OF SPEECH FRAMES 958
ESTIMATING A PLURALITY OF LINEAR PREDICTION COEFFICIENTS 960
PERFORM A WINDOWING ON THE PRE-EMPHASIZED PLURALITY OF SPEECH FRAMES 962
INVERSE FILTER THE PRE-PROCESSED SPEECH FRAMES TO OBTAIN THE RESIDUAL SIGNAL 964
B

FIG. 9D

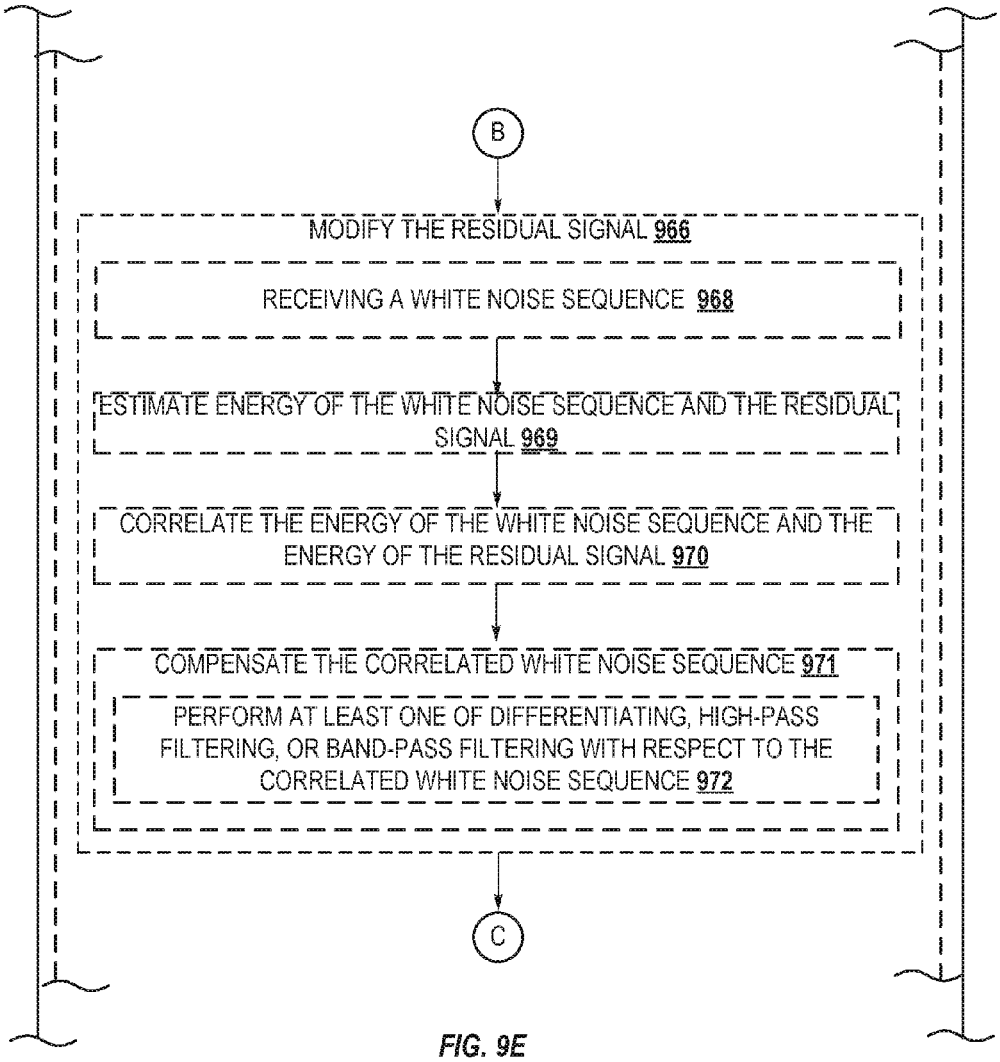

B
MODIFY THE RESIDUAL SIGNAL 966
RECEIVING A WHITE NOISE SEQUENCE 968
ESTIMATE ENERGY OF THE WHITE NOISE SEQUENCE AND THE RESIDUAL SIGNAL 969
CORRELATE THE ENERGY OF THE WHITE NOISE SEQUENCE AND THE ENERGY OF THE RESIDUAL SIGNAL 970
COMPENSATE THE CORRELATED WHITE NOISE SEQUENCE 971
PERFORM AT LEAST ONE OF DIFFERENTIATING, HIGH-PASS FILTERING, OR BAND-PASS FILTERING WITH RESPECT TO THE CORRELATED WHITE NOISE SEQUENCE 972
C

FIG. 9E

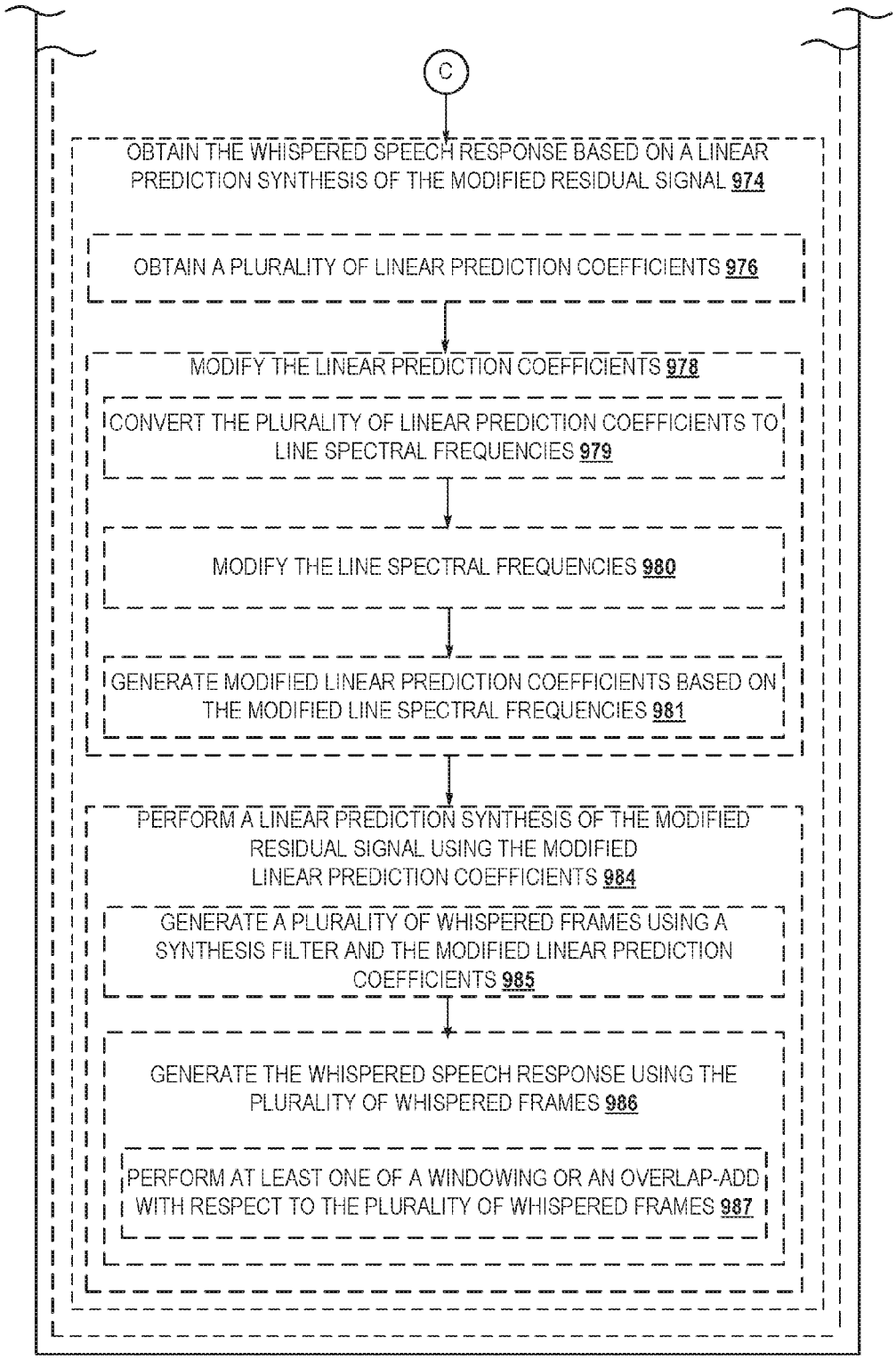

C
OBTAIN THE WHISPERED SPEECH RESPONSE BASED ON A LINEAR PREDICTION SYNTHESIS OF THE MODIFIED RESIDUAL SIGNAL 974
OBTAIN A PLURALITY OF LINEAR PREDICTION COEFFICIENTS 976
MODIFY THE LINEAR PREDICTION COEFFICIENTS 978
CONVERT THE PLURALITY OF LINEAR PREDICTION COEFFICIENTS TO LINE SPECTRAL FREQUENCIES 979
MODIFY THE LINE SPECTRAL FREQUENCIES 980
GENERATE MODIFIED LINEAR PREDICTION COEFFICIENTS BASED ON THE MODIFIED LINE SPECTRAL FREQUENCIES 981
PERFORM A LINEAR PREDICTION SYNTHESIS OF THE MODIFIED RESIDUAL SIGNAL USING THE MODIFIED LINEAR PREDICTION COEFFICIENTS 984
GENERATE A PLURALITY OF WHISPERED FRAMES USING A SYNTHESIS FILTER AND THE MODIFIED LINEAR PREDICTION COEFFICIENTS 985
GENERATE THE WHISPERED SPEECH RESPONSE USING THE PLURALITY OF WHISPERED FRAMES 986
PERFORM AT LEAST ONE OF A WINDOWING OR AN OVERLAP-ADD WITH RESPECT TO THE PLURALITY OF WHISPERED FRAMES 987

FIG. 9F

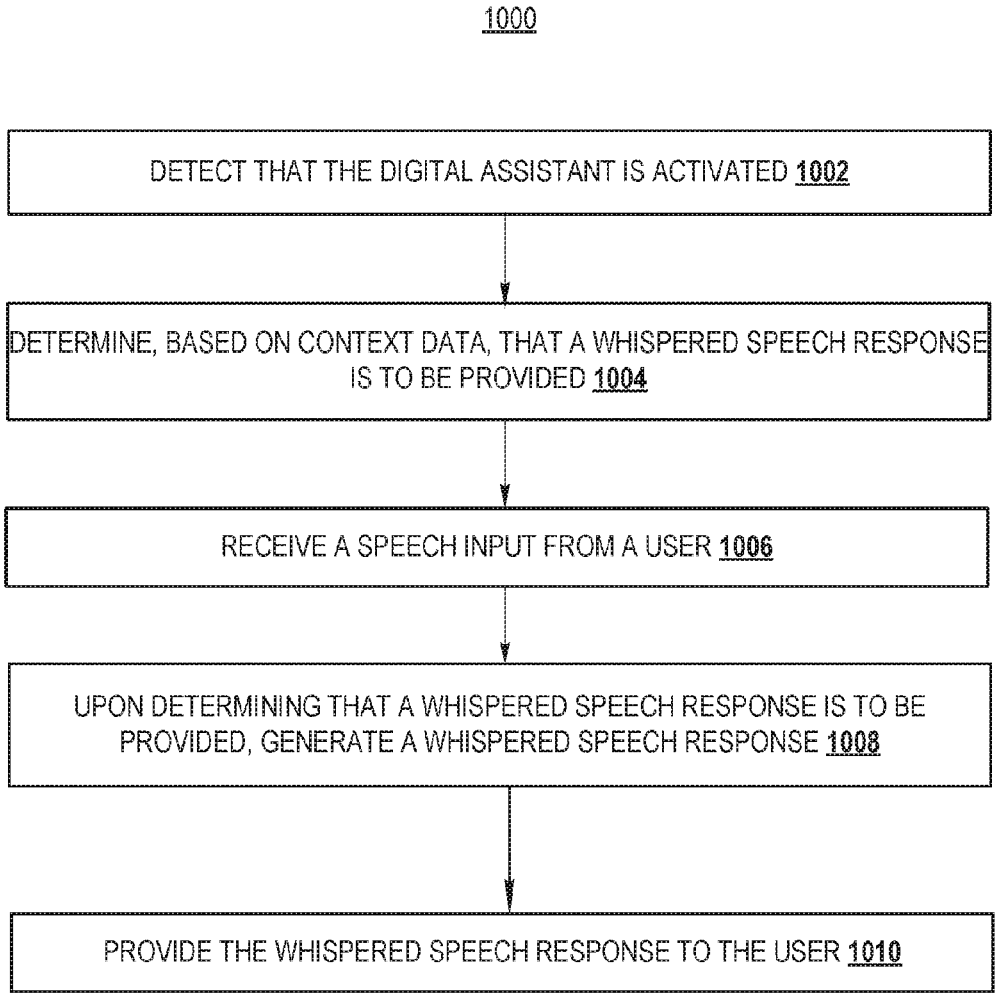

1000
DETECT THAT THE DIGITAL ASSISTANT IS ACTIVATED 1002
DETERMINE, BASED ON CONTEXT DATA, THAT A WHISPERED SPEECH RESPONSE IS TO BE PROVIDED 1004
RECEIVE A SPEECH INPUT FROM A USER 1006
UPON DETERMINING THAT A WHISPERED SPEECH RESPONSE IS TO BE PROVIDED, GENERATE A WHISPERED SPEECH RESPONSE 1008
PROVIDE THE WHISPERED SPEECH RESPONSE TO THE USER 1010

FIG. 10

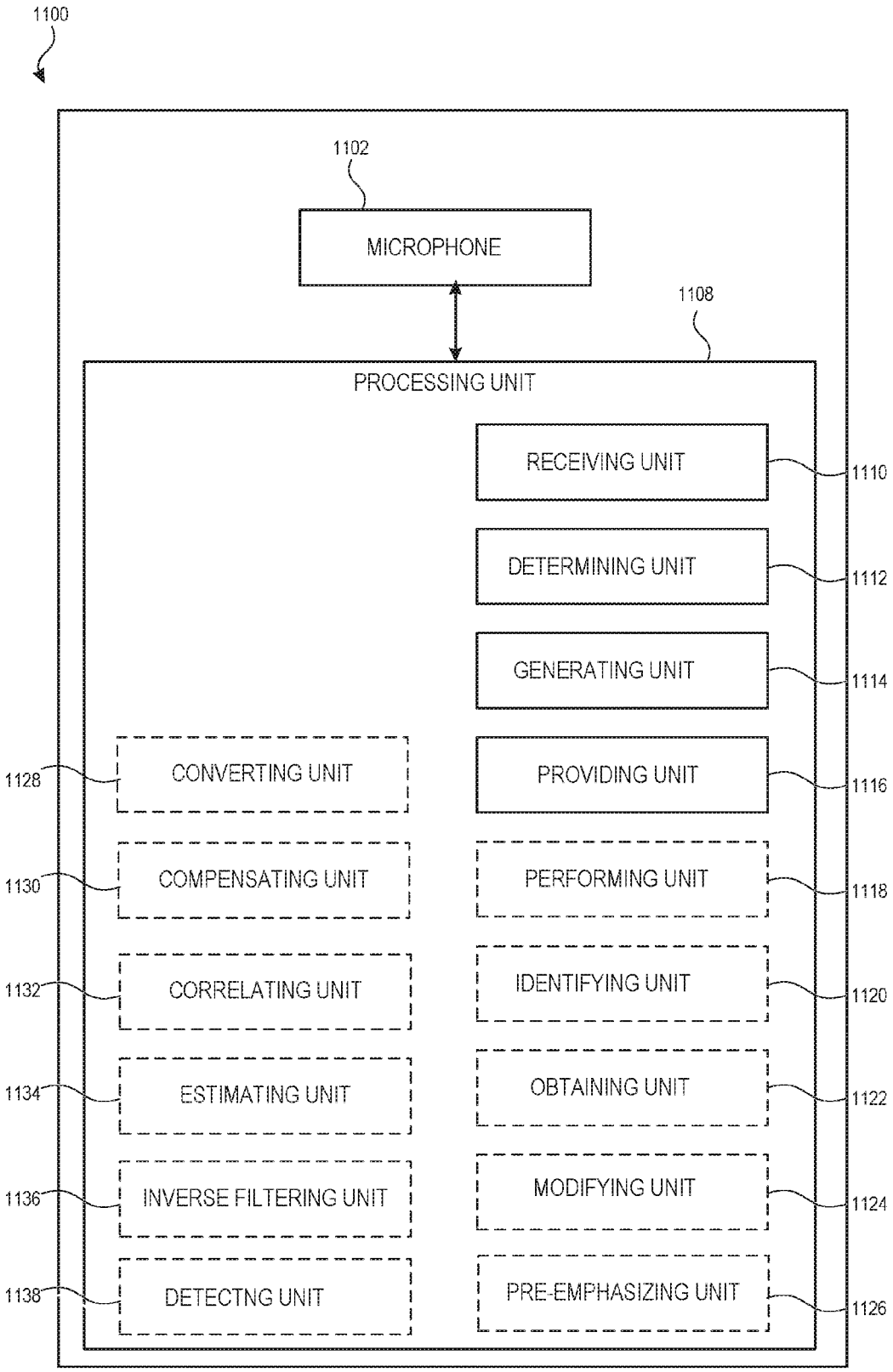

1100
1102
MICROPHONE
1108
PROCESSING UNIT
RECEIVING UNIT
1110
DETERMINING UNIT
1112
GENERATING UNIT
1114
1128
CONVERTING UNIT
PROVIDING UNIT
1116
1130
COMPENSATING UNIT
PERFORMING UNIT
1118
1132
CORRELATING UNIT
IDENTIFYING UNIT
1120
1134
ESTIMATING UNIT
OBTAINING UNIT
1122
1136
INVERSE FILTERING UNIT
MODIFYING UNIT
1124
1138
DETECTNG UNIT
PRE-EMPHASIZING UNIT
1126

FIG. 11

DIGITAL ASSISTANT PROVIDING WHISPERED SPEECH

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Ser. No. 62/348,705, filed on Jun. 10, 2016, entitled "DIGITAL ASSISTANT PROVIDING WHISPERED SPEECH," 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 is capable of detecting a whispered speech input and providing a whispered speech response.

BACKGROUND

[0003] Speech recognition devices are increasingly popular. A speech recognition device may be used under different circumstances or in different environments. For example, a user may ask the device a question while working at a cubicle with other co-workers surrounding the user. A user may also ask the device a question while attending a meeting in a conference room with other meeting participants. A user may also speak to the device while studying in a library where speaking loudly may be prohibited. Under some circumstances, the user may whisper speech to the device and/or desire that responses be whispered. It is thus important that a device recognizes the user's whispered speech, which may vary from normal speech, and provide a whispered speech response under these circumstances.

BRIEF SUMMARY

[0004] Some techniques for speech recognition detect a user's speech and provide a response without taking into consideration the circumstances associated with the user's speech. For example, a user may be attending a meeting or studying in a library. Under such circumstances, the user may not be able to speak in his or her regular voice. Rather, the user may whisper. Some existing speech recognition devices may be incapable of detecting the user's whispered speech, or suffer from misrecognition of the whispered speech. Nor are they capable of responding to the user using a whispered speech. For example, they may respond in a regular or loud voice, which may be undesired under certain circumstances where a whispered response is more appropriate. Providing a digital assistant that is capable of detecting whispered a speech input and providing a whispered speech response is thus important.

[0005] 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 speech input from a user. The method also includes determining, based on the speech input, that a whispered speech response is to be provided. The method further includes, upon determining that a whispered speech response is to be provided, generating a whispered speech response, and providing the whispered speech response to the user.

[0006] 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

[0007] 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.

[0008] FIG. **1** is a block diagram illustrating a system and environment for implementing a digital assistant according to various examples.

[0009] FIG. **2**A is a block diagram illustrating a portable multifunction device implementing the client-side portion of a digital assistant in accordance with some embodiments.

[0010] FIG. **2**B is a block diagram illustrating exemplary components for event handling according to various examples.

[0011] FIG. **3** illustrates a portable multifunction device implementing the client-side portion of a digital assistant according to various examples.

[0012] FIG. **4** is a block diagram of an exemplary multifunction device with a display and a touch-sensitive surface according to various examples.

[0013] FIG. **5**A illustrates an exemplary user interface for a menu of applications on a portable multifunction device according to various examples.

[0014] FIG. **5**B illustrates an exemplary user interface for a multifunction device with a touch-sensitive surface that is separate from the display according to various examples.

[0015] FIG. **6**A illustrates a personal electronic device according to various examples.

[0016] FIG. **6**B is a block diagram illustrating a personal electronic device according to various examples.

[0017] FIG. **7**A is a block diagram illustrating a digital assistant system or a server portion thereof according to various examples.

[0018] FIG. **7**B illustrates the functions of the digital assistant shown in FIG. **7**A according to various examples.

[0019] FIG. **7**C illustrates a portion of an ontology according to various examples.

[0020] FIG. **8**A illustrates a block diagram of functions of a digital assistant according to various examples.

[0021] FIG. **8**B illustrates an exemplary spectrum of a non-whispered speech input according to various examples.

[0022] FIG. **8**C illustrates an exemplary spectrum of a whispered speech input according to various examples.

[0023] FIG. **8**D illustrates a block diagram of the functions of a whisper conversion module according to various examples.

[0024] FIGS. **9**A-**9**F illustrate a flow diagram of an exemplary process for operating a digital assistant according to various examples.

[0025] FIG. **10** illustrates a flow diagram of an exemplary process for operating a digital assistant according to various examples.

[0026] FIG. **11** illustrates a block diagram of an electronic device according to various examples.

DETAILED DESCRIPTION

[0027] 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.

[0028] Techniques for providing whispered speech response are desirable. As described herein, techniques for determining that a whispered speech response is to be provided are desired for various purposes, such as avoiding disturbing others and protecting the user's privacy. Such techniques are advantageous because they allow the user to whisper to a digital assistant and receive a whispered speech response from the digital assistant. Further, such techniques for providing whispered speech response can allow the user to use the digital assistant under certain circumstances without having to change the user's current physical location (e.g., step outside to use the digital assistant).

[0029] 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 slope could be termed a second slope and, similarly, a second slope could be termed a first slope, without departing from the scope of the various described examples. The first slope and the second slope can both be slopes and, in some cases, can be separate and different slopes.

[0030] 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.

[0031] 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

[0032] 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.

[0033] 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.

[0034] 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**.

[0035] 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.

[0036] 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. **2**A), 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. **6**A-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, Calif. Other examples of portable multifunction devices can include, without limitation, laptop or tablet computers. Further, in some examples, user device **104** can be a non-portable multifunctional device. In particular, user device **104** can be a desktop computer, a game console, a television, or a television set-top box. In some examples, user device **104** can include a touch-sensitive surface (e.g., touch screen displays and/or touchpads). Further, user device **104** can optionally include one or more other physical user-interface devices, such as a physical keyboard, a mouse, and/or a joystick. Various examples of electronic devices, such as multifunctional devices, are described below in greater detail.

[0037] 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.

[0038] 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**.

[0039] 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. **2**A, **4**, and **6**A-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**.

[0040] 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**.

[0041] 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

[0042] Attention is now directed toward embodiments of electronic devices for implementing the client-side portion of a digital assistant. FIG. **2**A 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**.

[0043] 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).

[0044] 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.

[0045] 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. **2**A 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.

[0046] 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**.

[0047] 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.

[0048] 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.

[0049] 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.

[0050] 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).

[0051] 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**).

[0052] 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 Ser. No. 11/322,549, "Unlocking a Device by Performing Gestures on an Unlock Image," filed Dec. 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.

[0053] 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.

[0054] 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.

[0055] 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, Calif.

[0056] 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. Pat. No.6,323,846 (Westerman et al.), U.S. Pat. No. 6,570,557 (Westerman et al.), and/or U.S. Pat. No. 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.

[0057] A touch-sensitive display in some embodiments of touch screen **212** may be as described in the following applications: (1) U.S. patent application Ser. No. 11/381,313, "Multipoint Touch Surface Controller," filed May 2, 2006; (2) U.S. patent application Ser. No. 10/840,862, "Multipoint Touchscreen," filed May 6, 2004; (3) U.S. patent application Ser. No. 10/903,964, "Gestures For Touch Sensitive Input Devices," filed Jul. 30, 2004; (4) U.S. patent application Ser. No. 11/048,264, "Gestures For Touch Sensitive Input Devices," filed Jan. 31, 2005; (5) U.S. patent application Ser. No. 11/038,590, "Mode-Based Graphical User Interfaces For Touch Sensitive Input Devices," filed Jan. 18, 2005; (6) U.S. patent application Ser. No. 11/228,758, "Virtual Input Device Placement On A Touch Screen User Interface," filed Sep. 16, 2005; (7) U.S. patent application Ser. No. 11/228,700, "Operation Of A Computer With A Touch Screen Interface," filed Sep. 16, 2005; (8) U.S. patent application Ser. No. 11/228,737, "Activating Virtual Keys Of A Touch-Screen Virtual Keyboard," filed Sep. 16, 2005; and (9) U.S. patent application Ser. No. 11/367,749, "Multi-Functional Hand-Held Device," filed Mar. 3, 2006. All of these applications are incorporated by reference herein in their entirety.

[0058] 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.

[0059] 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.

[0060] Device **200** also includes power system **262** for powering the various components. Power system **262** may include a power management system, one or more power sources (e.g., battery, alternating current (AC)), a recharging system, a power failure detection circuit, a power converter or inverter, a power status indicator (e.g., a light-emitting diode (LED)) and any other components associated with the generation, management and distribution of power in portable devices.

[0061] Device **200** may also include one or more optical sensors **264**. FIG. **2**A 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.

[0062] Device **200** optionally also includes one or more contact intensity sensors **265**. FIG. **2**A 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**.

[0063] Device **200** may also include one or more proximity sensors **266**. FIG. **2**A 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 Ser. No. 11/241,839, "Proximity Detector In Handheld Device"; Ser. No. 11/240,788, "Proximity Detector In Handheld Device"; Ser. No. 11/620,702, "Using Ambient Light Sensor To Augment Proximity Sensor Output"; Ser. No. 11/586,862, "Automated Response To And Sensing Of User Activity In Portable Devices"; and Ser. No. 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).

[0064] Device **200** optionally also includes one or more tactile output generators **267**. FIG. **2**A 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**.

[0065] Device **200** may also include one or more accelerometers **268**. FIG. **2**A 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**.

[0066] 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. **2**A) or **470** (FIG. **4**) stores device/global internal state **257**, as shown in FIGS. **2**A 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.

[0067] 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.

[0068] 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.

[0069] 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.

[0070] 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).

[0071] 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.

[0072] 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.

[0073] 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**.

[0074] 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**.

[0075] 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).

[0076] 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).

[0077] Digital assistant client module **229** can include various client-side digital assistant instructions to provide the client-side functionalities of the digital assistant. For example, digital assistant client module **229** can be capable of accepting voice input (e.g., speech input), text input, touch input, and/or gestural input through various user interfaces (e.g., microphone **213**, accelerometer(s) **268**, touch-sensitive display system **212**, optical sensor(s) **229**, other input control devices **216**, etc.) of portable multifunction device **200**. Digital assistant client module **229** can also be capable of providing output in audio (e.g., speech output), visual, and/or tactile forms through various output interfaces (e.g., speaker **211**, touch-sensitive display system **212**, tactile output generator(s) **267**, etc.) of portable multifunction device **200**. For example, output can be provided as voice, sound, alerts, text messages, menus, graphics, videos, animations, vibrations, and/or combinations of two or more of the above. During operation, digital assistant client module **229** can communicate with DA server **106** using RF circuitry **208**.

[0078] 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.

[0079] 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.

[0080] 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.

[0081] 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.

[0082] A more detailed description of a digital assistant is described below with reference to FIGS. **7**A-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.

[0083] Applications **236** may include the following modules (or sets of instructions), or a subset or superset thereof:

- **[0084]** Contacts module **237** (sometimes called an address book or contact list);
- **[0085]** Telephone module **238**;
- **[0086]** Video conference module **239**;
- **[0087]** Email client module **240**;
- **[0088]** Instant messaging (IM) module **241**;
- **[0089]** Workout support module **242**;
- **[0090]** Camera module **243** for still and/or video images;
- **[0091]** Image management module **244**;
- **[0092]** Video player module;
- **[0093]** Music player module;
- **[0094]** Browser module **247**;
- **[0095]** Calendar module **248**;
- **[0096]** 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**;
- **[0097]** Widget creator module **250** for making user-created widgets **249-6**;
- **[0098]** Search module **251**;

[0099] Video and music player module **252**, which merges video player module and music player module;

[0100] Notes module **253**;

[0101] Map module **254**; and/or

[0102] Online video module **255**.

[0103] 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.

[0104] 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.

[0105] 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.

[0106] 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.

[0107] 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**.

[0108] In conjunction with RF circuitry **208**, touch screen **212**, display controller **256**, contact/motion module **230**, graphics module **232**, and text input module **234**, the instant messaging module **241** includes executable instructions to enter a sequence of characters corresponding to an instant message, to modify previously entered characters, to transmit a respective instant message (for example, using a Short Message Service (SMS) or Multimedia Message Service (MMS) protocol for telephony-based instant messages or using XMPP, SIMPLE, or IMPS for Internet-based instant messages), to receive instant messages, and to view received instant messages. In some embodiments, transmitted and/or received instant messages may include graphics, photos, audio files, video files and/or other attachments as are supported in an MMS and/or an Enhanced Messaging Service (EMS). As used herein, "instant messaging" refers to both telephony-based messages (e.g., messages sent using SMS or MMS) and Internet-based messages (e.g., messages sent using XMPP, SIMPLE, or IMPS).

[0109] 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.

[0110] 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**.

[0111] 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.

[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**, 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.

[0113] 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.

[0114] 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).

[0115] 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).

[0116] 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.

[0117] 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.).

[0118] 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.

[0119] 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.

[0120] 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 Jun. 20, 2007, and U.S. patent application Ser. No. 11/968,067, "Portable Multifunction Device, Method, and Graphical User Interface for Playing Online Videos," filed Dec. 31, 2007, the contents of which are hereby incorporated by reference in their entirety.

[0121] 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. **2**A). 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.

[0122] 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.

[0123] 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.

[0124] FIG. **2**B is a block diagram illustrating exemplary components for event handling in accordance with some embodiments. In some embodiments, memory **202** (FIG. **2**A) 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**).

[0125] 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.

[0126] 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.

[0127] 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.

[0128] 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).

[0129] In some embodiments, event sorter **270** also includes a hit view determination module **272** and/or an active event recognizer determination module **273**.

[0130] 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.

[0131] 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.

[0132] 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.

[0133] 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.

[0134] 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**.

[0135] 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**.

[0136] 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**.

[0137] 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).

[0138] 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.

[0139] 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**.

[0140] 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.

[0141] 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.

[0142] 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.

[0143] 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.

[0144] 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.

[0145] 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.

[0146] 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.

[0147] 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.

[0148] 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.

[0149] 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.

[0150] 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**.

[0151] 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**.

[0152] 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. **2**A), 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. **2**A). 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. **2**A), 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. **2**A) optionally does not store these modules.

[0153] 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.

[0154] Attention is now directed towards embodiments of user interfaces that may be implemented on, for example, portable multifunction device **200**.

[0155] FIG. **5**A 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:

[0156] Signal strength indicator(s) **502** for wireless communication(s), such as cellular and Wi-Fi signals;

- **[0157]** Time **504**;
- **[0158]** Bluetooth indicator **505**;
- **[0159]** Battery status indicator **506**;
- **[0160]** Tray **508** with icons for frequently used applications, such as:
  - **[0161]** Icon **516** for telephone module **238**, labeled "Phone," which optionally includes an indicator **514** of the number of missed calls or voicemail messages;
  - **[0162]** Icon **518** for email client module **240**, labeled "Mail," which optionally includes an indicator **510** of the number of unread emails;
  - **[0163]** Icon **520** for browser module **247**, labeled "Browser;" and
  - **[0164]** Icon **522** for video and music player module **252**, also referred to as iPod (trademark of Apple Inc.) module **252**, labeled "iPod;" and
- **[0165]** Icons for other applications, such as:
  - **[0166]** Icon **524** for IM module **241**, labeled "Messages;"
  - **[0167]** Icon **526** for calendar module **248**, labeled "Calendar;"
  - **[0168]** Icon **528** for image management module **244**, labeled "Photos;"
  - **[0169]** Icon **530** for camera module **243**, labeled "Camera;"
  - **[0170]** Icon **532** for online video module **255**, labeled "Online Video;"
  - **[0171]** Icon **534** for stocks widget **249-2**, labeled "Stocks;"
  - **[0172]** Icon **536** for map module **254**, labeled "Maps;"
  - **[0173]** Icon **538** for weather widget **249-1**, labeled "Weather;"

[0174] Icon **540** for alarm clock widget **249-4**, labeled "Clock;"

[0175] Icon **542** for workout support module **242**, labeled "Workout Support;"

[0176] Icon **544** for notes module **253**, labeled "Notes;" and

[0177] Icon **546** for a settings application or module, labeled "Settings," which provides access to settings for device **200** and its various applications **236**.

[0178] It should be noted that the icon labels illustrated in FIG. **5**A 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.

[0179] FIG. **5**B 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**.

[0180] 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. **5**B. In some embodiments, the touch-sensitive surface (e.g., **551** in FIG. **5**B) has a primary axis (e.g., **552** in FIG. **5**B) that corresponds to a primary axis (e.g., **553** in FIG. **5**B) on the display (e.g., **550**). In accordance with these embodiments, the device detects contacts (e.g., **560** and **562** in FIG. **5**B) with the touch-sensitive surface **551** at locations that correspond to respective locations on the display (e.g., in FIG. **5**B, **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. **5**B) are used by the device to manipulate the user interface on the display (e.g., **550** in FIG. **5**B) 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.

[0181] 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.

[0182] FIG. **6**A 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. **2**A-**4**B). 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**.

[0183] 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 Nov. 11, 2013, each of which is hereby incorporated by reference in their entirety.

[0184] 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.

[0185] FIG. **6**B depicts exemplary personal electronic device **600**. In some embodiments, device **600** can include some or all of the components described with respect to FIGS. **2**A, **2**B, 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.

[0186] 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**.

[0187] 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. **12**A-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. **6**B, but can include other or additional components in multiple configurations.

[0188] 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.

[0189] 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. **5**B) 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. **2**A or touch screen **212** in FIG. **5**A) 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).

[0190] 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.

[0191] 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.

[0192] 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.

[0193] 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.

[0194] 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).

[0195] 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).

[0196] 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

[0197] FIG. **7**A 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. **7**A 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.

[0198] 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**.

[0199] 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).

[0200] In some examples, I/O interface **706** can couple input/output devices **716** of digital assistant system **700**, such as displays, keyboards, touch screens, and microphones, to user interface module **722**. I/O interface **706**, in conjunction with user interface module **722**, can receive user inputs (e.g., voice input, keyboard inputs, touch inputs, etc.) and processes them accordingly. In some examples, e.g., when the digital assistant is implemented on a standalone user device, digital assistant system **700** can include any of the components and I/O communication interfaces described with respect to devices **200**, **400**, or **600** in FIGS. **2**A, **4**, **6**A-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**).

[0201] In some examples, the network communications interface **708** can include wired communication port(s) **712** and/or wireless transmission and reception circuitry **714**. The wired communication port(s) can receive and send communication signals via one or more wired interfaces, e.g., Ethernet, Universal Serial Bus (USB), FIREWIRE, etc. The wireless circuitry **714** can receive and send RF signals and/or optical signals from/to communications networks and other communications devices. The wireless communications can use any of a plurality of communications standards, protocols, and technologies, such as GSM, EDGE, CDMA, TDMA, Bluetooth, Wi-Fi, VoIP, Wi-MAX, or any other suitable communication protocol. Network communications interface **708** can enable communication between digital assistant system **700** with networks, such as the Internet, an intranet, and/or a wireless network, such as a cellular telephone network, a wireless local area network (LAN), and/or a metropolitan area network (MAN), and other devices.

[0202] 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 reads/writes from/to the data structures.

[0203] 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.

[0204] 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. **2**A, **4**, **6**A-B, respectively. Communications module **720** can also include various components for handling data received by wireless circuitry **714** and/or wired communications port **712**.

[0205] 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.).

[0206] 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.

[0207] 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.

[0208] 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.

[0209] In some examples, as shown in FIG. **7**B, I/O processing module **728** can interact with the user through I/O devices **716** in FIG. **7**A or with a user device (e.g., devices **104**, **200**, **400**, or **600**) through network communications interface **708** in FIG. **7**A to obtain user input (e.g., a speech input) and to provide responses (e.g., as speech outputs) to the user input. I/O processing module **728** can optionally obtain contextual information associated with the user input from the user device, along with or shortly after the receipt of the user input. The contextual information can include user-specific data, vocabulary, and/or preferences relevant to the user input. In some examples, the contextual information also includes software and hardware states of the user device at the time the user request is received, and/or information related to the surrounding environment of the user at the time that the user request was received. In some examples, I/O processing module **728** can also send follow-up questions to, and receive answers from, the user regarding the user request. When a user request is received by I/O processing module **728** and the user request can include speech input, I/O processing module **728** can forward the speech input to STT processing module **730** (or speech recognizer) for speech-to-text conversions.

[0210] 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.

[0211] More details on the speech-to-text processing are described in U.S. Utility Application Ser. No. 13/236,942 for "Consolidating Speech Recognition Results," filed on Sep. 20, 2011, the entire disclosure of which is incorporated herein by reference.

[0212] 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/ˈtəˈmeɪɾoʊ /and/ˈtəˈmɑtoʊ /. 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.

[0213] In some examples, the candidate pronunciations can be ranked based on the commonness of the candidate pronunciation. For example, the candidate pronunciation//ˈtəˈmeɪɾoʊ / can be ranked higher than/ˈtəˈmɑtoʊ /, 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/ˈtəˈmeɪɾoʊ /can be associated with the United States, whereas the candidate pronunciation/ˈtəˈmɑtoʊ /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/ˈtəˈmeɪɾoʊ /(associated with the United States) can be ranked higher than the candidate pronunciation/ˈtəˈmɑtoʊ /(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).

[0214] 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/ˈtəˈmeɪɾoʊ /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."

[0215] 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/

ˈtəˈmeɪroʊ /corresponds to the word "tomato," even if that particular sequence of phonemes is not one of the candidate sequence of phonemes for that word.

[0216] 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.

[0217] 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.

[0218] 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.

[0219] 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. **7**C, 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).

[0220] 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. **7**C, 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**.

[0221] 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. **7**C 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**.

[0222] While FIG. **7**C 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."

[0223] 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**.

[0224] 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."

[0225] 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.

[0226] 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.

[0227] 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.

[0228] Other details of searching an ontology based on a token string is described in U.S. Utility Application Ser. No. 12/341,743 for "Method and Apparatus for Searching Using An Active Ontology," filed Dec. 22, 2008, the entire disclosure of which is incorporated herein by reference.

[0229] 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.

[0230] 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.

[0231] 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 receives and processes the user responses. The questions can be provided to and answers can

be received from the users through I/O processing module **728**. In some examples, dialogue flow processing module **734** can present dialogue output to the user via audio and/or visual output, and receives input from the user via spoken or physical (e.g., clicking) responses. Continuing with the example above, when task flow processing module **736** invokes dialogue flow processing module **734** to determine the "party size" and "date" information for the structured query associated with the domain "restaurant reservation," dialogue flow processing module **734** can generate questions such as "For how many people?" and "On which day?" to pass to the user. Once answers are received from the user, dialogue flow processing module **734** can then populate the structured query with the missing information, or pass the information to task flow processing module **736** to complete the missing information from the structured query.

[0232] 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=Mar. 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.

[0233] 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.

[0234] 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.

[0235] 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.

[0236] 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.

[0237] 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.

[0238] Additional details on digital assistants can be found in the U.S. Utility Application Ser. No. 12/987,982, entitled "Intelligent Automated Assistant," filed Jan. 10, 2011, and U.S. Utility Application Ser. No. 13/251,088, entitled "Generating and Processing Task Items That Represent Tasks to Perform," filed Sep. 30, 2011, the entire disclosures of which are incorporated herein by reference.

4. Exemplary Functions of a Digital Assistant

[0239] FIG. **8**A illustrates a block diagram of functions of a digital assistant **800** (or the server portion of a digital assistant). In some examples, digital assistant **800** can be implemented using digital assistant module **726**. Digital assistant **800** can include one or more modules, models,

applications, vocabularies, and user data similar to those of digital assistant module **726**. For example, digital assistant **800** can include the following sub-modules, or a subset or superset thereof: input/output processing module **828** and speech synthesis module **840**. Input/output processing module **828** and speech synthesis module **840** can be implemented in a manner similar to input/output processing module **728** and speech synthesis module **740** as illustrated in FIG. **7**B, respectively. In some examples, digital assistant **800** can also include a speech-to-text (STT) processing module, a natural language processing module, a dialogue flow processing module, a task flow processing module, a service processing module, and a speech synthesis module. These modules can also be implemented similar to that of the corresponding modules as illustrated in FIG. **7**B, and therefore are not shown and not repeatedly described.

[0240] In some embodiments, digital assistant **800** can include a whispered speech determination module **820** and a whisper conversion module **810**. Whispered speech determination module **820** can determine whether a whispered speech response is to be provided, for example, based on one or both of a speech input and context data. As shown in FIG. **8**A, in some examples, a user **830** provides a speech input to digital assistant **800**. The speech input can include an informational request, a request to perform a task, or a combination of the two. An informational request seeks an informational answer. An informational answer provides the information the user seeks and may relate to any subject matter such as weather, traffic, news, finance, sports, or the like. For example, user **830** may ask digital assistant **800** a question such as "What time is it in Paris now?" In response, digital assistant **800** may answer "It is 2 p.m. in Paris right now."

[0241] In some examples, the speech input can include a request to perform a task. For example, the speech input may include "Set a reminder for the meeting next Tuesday with John at 2 p.m.." In response, digital assistant **800** may generate a reminder in the user's calendar and acknowledge the request by outputting a speech response such as "Reminder for the meeting next Tuesday with John at 2 p.m. is set." In some examples, during the performance of a requested task, digital assistant **800** can interact with user **830** in a continuous dialogue involving multiple exchanges of information over an extended period of time. There are numerous other ways of interacting with digital assistant **800** to request information or request the performance various tasks. In addition to providing speech responses and taking programmed actions, digital assistant **800** can also provide responses in other visual or audio forms, e.g., as text, alerts, music, videos, animations, etc.

[0242] As shown in FIG. **8**A, in some examples, in response to receiving a speech input from user **830**, whispered speech determination module **820** can determine whether the speech input includes a whispered speech input. Under certain circumstances, user **830** may whisper to digital assistant **800**. For example, user **830** may be attending a meeting with other meeting participants in the conference room, or studying in a quiet library. Under such circumstances, user **830** may whisper to digital assistant **800** to avoid disturbing other people and/or to protect the user's privacy.

[0243] Whispered speech determination module **820** can determine whether the speech input includes a whispered speech input based on one or more differences between a whispered speech input and a non-whispered speech input. A whispered speech input may be different from a non-whispered speech input in various aspects. For example, to generate a non-whispered speech input, vocal folds of user **830** vibrate and produce a periodic excitation signal. The periodic excitation signal results in a non-whispered phonation (e.g., phonation in regular voice). To generate a whispered speech input, user **830** produces an excitation signal by adducting the vocal folds so that the air passing through the glottis becomes turbulent, but not enough to start a vibration. This excitation signal may result in whispered phonation.

[0244] As a result of the different excitation signals associated with a non-whispered speech input and a whispered speech input, the non-whispered speech input and the whispered speech input can have differing spectrums. FIG. **8**B illustrates an exemplary spectrum **832** of a non-whispered speech input according to various examples. FIG. **8**C illustrates an exemplary spectrum **834** of a whispered speech input according to various examples. In both spectrum **832** and spectrum **834**, the horizontal axis represents frequency and the vertical axis represents amplitude. In some examples, a whispered speech input can have one or more spectrum characteristics that are different from the corresponding spectrum characteristics of a non-whispered speech input. For example, spectrum **832** of a non-whispered speech input may have an amplitude that is less than the corresponding amplitude of spectrum **834** of a whispered speech input below a pre-determined frequency (e.g., 800 Hz). Spectrum **832** of a non-whispered speech input may also have an energy that is less than the corresponding energy of spectrum **834** of a whispered speech input below a pre-determined frequency (e.g., 800 Hz). In some examples, spectrum **832** of a non-whispered speech input may have a volume that is less than the corresponding volume of spectrum **834** of a whispered speech input. And spectrum **832** of a non-whispered speech input may have a slope or tilt that is shifted by a threshold slope percentage with respect to the slope of spectrum **834** of a whispered speech input. For example, as shown in FIGS. **8**B and **8**C, in a pre-determined frequency range (e.g., below 800 Hz), an absolute value of the slope of spectrum **832** of a non-whispered speech input can be greater than the absolute value of the slope of spectrum **834** of a whispered speech input by a threshold slope percentage.

[0245] As described and shown in FIG. **8**A, in response to receiving a speech input from user **830**, whispered speech determination module **820** can determine whether the speech input includes a whispered speech input. In some examples, whispered speech determination module **820** can make such determination based on one or more spectrum characteristics, such as the amplitude, the energy, the volume, the slope, or a combination thereof. For example, whispered speech determination module **820** can obtain the spectrum of the speech input and determine one or more spectrum characteristics such as the amplitude, the energy, the volume, and the slope. Whispered speech determination module **820** can further determine whether one or more of the amplitude, the energy, the volume, and the slope associated with the spectrum of the speech input satisfies a corresponding threshold condition. For example, if the amplitude of the spectrum of the speech input is smaller than a threshold amplitude below pre-determined frequency (e.g., 800 Hz), whispered speech determination module **820** may determine

that the speech input includes a whispered speech input. As another example, whispered speech determination module **820** can also obtain both the amplitude and the slope of the spectrum of the speech input in a pre-determined frequency range (e.g., below 800 Hz). If both the amplitude and the slope satisfy the corresponding threshold conditions (e.g., the amplitude is less than a threshold amplitude and the absolute value of the slope is smaller than a threshold slope), whispered speech determination module **820** may determine that the speech input includes a whispered speech input. In some embodiments, using more than one spectrum characteristic may increase the accuracy in determining whether the speech input includes a whispered speech input.

[0246] In some embodiments, whispered speech determination module **820** can determine whether a speech input includes a whispered speech input using one or more features of the speech input. A feature of the speech input can represent one or more spectrum characteristics associated with the spectrum of the speech input. For example, a feature may include a mel-frequency cepstrum coefficient (MFCC). A plurality of MFCCs can represent a mel-frequency cepstrum (MFC). An MFC can be a representation of a short-term power spectrum of the speech input.

[0247] In some examples, whispered speech determination module **820** can determine whether a speech input includes a whispered speech input using two features associated with the spectrum of the speech input. For example, whispered speech determination module **820** can obtain the spectrum of the speech input and determine one or more of spectrum characteristics (e.g., an amplitude, an energy, a volume, and a slope). Based on the one or more spectrum characteristics, whispered speech determination module **820** can determine a first feature and a second feature. For example, a first feature can include a first MFCC (e.g., MFCC**0**) representing an energy or an amplitude associated with the spectrum of the speech input. A second feature can include a second MFCC (e.g., MFCC**1**) representing a slope or tilt associated with the spectrum of the speech input. In some examples, whispered speech determination module **820** can obtain a whisper score based on the first feature to the second feature. In some examples, a whisper score can be determined using MFCC**0** and MFCC**1** associated with each frame of the speech input. For example, as shown below in formula (1), a whisper score can be determined as a weighted mean of MFCC**1** values using MFCC**0** values as weights.

$$\text{whisper score}=\text{sum_over_frames}(C_{0t}*C_{1t})/\text{sum_over_frames}(C_{0t}) \quad (1)$$

In the above formula (1), t represents the index of a frame in a speech input, C_{0t} represents MFCC**0** at frame t, and C_{1t} represents MFCC**1** at frame t. The whisper score provides a measure of spectral tilt that is weighted more heavily for higher amplitude portions of the speech input. The higher amplitude portions of the speech input may likely correspond to a human speech rather than background noise. In some examples, if the whisper score satisfies a score threshold, whispered speech determination module **820** can determine that the speech input includes a whispered speech input. In some examples, the smaller the whisper score, the more likely that the speech input includes a whispered speech input; and the higher the whisper score, the more likely that the speech input includes a non-whispered speech input. For example, whispered speech determination module **820** can obtain a whisper score based on MFCC**0** to MFCC**1**. If the whisper score of MFCC**0** to MFCC**1** is less than a pre-determined score threshold, whispered speech determination module **820** can determine that the speech input includes a whispered speech input.

[0248] In some embodiments, the ratio of a first feature to a second feature (or the second feature to the first feature) can be adjusted. For example, the ratio can be adjusted in accordance with the manner in which the user speaks. The manner in which a user speaks may vary from person to person, or even with the same user. For example, a user may naturally speak with a softer voice or lower energy than that of other users. A user may have speech disabilities that prevent him or her from speaking loudly. As a result, the user may sound as though he or she is whispering even though he or she is not (e.g., the spectrum of a non-whispered speech of the user may have a smaller amplitude, volume, or energy). The opposite may also occur. For example, a user may naturally speak loudly, and he or she may sounds as though he or she is speaking in a normal voice even though he or she is whispering (e.g., the spectrum of a whispered speech of the user may have a larger amplitude, volume, or energy). Accordingly, in some examples, the ratio of the first feature to the second feature (e.g., MFCC**0**/MFCC**1**) can be adjusted in accordance with the manner in which the user speaks. The adjustment may increase the accuracy of determination whether the speech input includes a whispered speech input. In some examples, the adjustment may be performed automatically and dynamically using, for example, machine learning techniques.

[0249] In some embodiments, whispered speech determination module **820** can determine whether a speech input includes a whispered speech input using a far-field acoustic model. A far-field acoustic model enables the detection of speech when the user speaks at an extended distance from digital assistant **800** (e.g., at a distance beyond a threshold distance). In some examples, the user's speech input in a far field can decay at a pre-determined rate (e.g., 6 dB) each time the distance from the user is doubled. When digital assistant **800** is in the far field, the user's speech may appear to be soft, and the spectrum of the speech input that digital assistant **800** receives may have a low volume, amplitude, and/or energy. In addition, in a far field, the speech input that digital assistant **800** receives may include echoes. In some examples, a whispered speech input may have characteristics (e.g., lower volume, energy, and/or amplitude) similar to those of the speech input that digital assistant **800** receives from a far-field user speech. Thus, in some examples, using a far-field acoustic model can increase the accuracy of whispered speech input detection. In some embodiments, whispered speech determination module **820** can switch to a far-field acoustic model after or when it determines that the speech input includes a whispered speech input. By switching to a far-field acoustic model, whispered speech determination module **820** can improve the accuracy of recognizing the subsequent user speech inputs.

[0250] As described above, whispered speech determination module **820** can determine whether a speech input includes a whispered speech input using one or more features of the speech input. It is appreciated that whispered speech determination module **820** can determine whether a speech input includes a whispered speech input using any desired classification methods and/or spectral estimation methods. For example, a neural network classifier can be used to determine whether a speech input includes a whis-

pered speech input. In some examples, it is appreciated that the features may include mel-cepstral coefficients that are used in speech synthesis.

[0251] As shown in FIG. **8**A, in some embodiments, whispered speech determination module **820** can determine whether a whispered speech response is expected based on context data. As described, context data (also referred to as contextual information) can include user-specific data, vocabulary, and/or preferences relevant to the user input. For example, the context data can include the user's calendar, the current time, the user's location data, and/or the user's preferences of operation of digital assistant **800**. For example, the user's current location data may indicate that the user is currently physically located inside a library. In some examples, the context data also includes the 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.

[0252] In some embodiments, whispered speech determination module **820** can obtain context data from, for example, the user device (e.g., the user's smartphone) and/or one or more devices communicatively connected to the user device (e.g., the user's laptop computer). After obtaining the context data, whispered speech determination module **820** can determine whether the context data satisfies one or more conditions for providing a whispered speech response. For example, if the user's location data indicates the user is located in a library for more than a threshold time period (e.g., 10 minutes), whispered speech determination module **820** can determine that a whispered speech response is expected. As another example, if the user's calendar data indicates that the user is currently attending a meeting with other meeting participants, whispered speech determination module **820** can determine that a whispered speech response is expected.

[0253] In some examples, whispered speech determination module **820** can determine whether a whispered speech response is to be provided based on the speech input only, based on the context data only, or based on both. As described above, if whispered speech determination module **820** determines that a speech input includes a whispered speech input, digital assistant **800** can determine that a whispered speech response is to be provided. In some examples, if whispered speech determination module **820** determines that the context data indicates that a whispered speech response is expected, digital assistant **800** can determine that a whispered speech response is to be provided, regardless of whether the speech input includes a whispered speech input. For example, as shown in FIG. **8**A, user **830** may provide a non-whispered speech input (e.g., in a regular loud voice) such as "Call John Doe." Whispered speech determination module **820** determines that the context data indicates that a whispered speech response is expected (e.g., user **830** is currently located in a library). As a result, digital assistant **800** can place the call to John Doe, and determine that a whispered speech response (e.g., "OK, calling John Doe") is to be provided.

[0254] In some embodiments, if whispered speech determination module **820** determines that the speech input includes a whispered speech input and/or that the context data indicates a whispered speech response is expected, it can determine that a whispered speech response is to be provided. As described above, a speech input can include an informational request, a request to perform a task, or a combination thereof. If the speech input includes an informational request, a whispered speech response can be provided to answer the informational request. For example, user **830** may whisper, "What is the time now in Paris?" In response, digital assistant **800** can generate and provide a whispered speech response such as "The current time in Paris is 2 p.m." If the speech input includes a request to perform a task, a whispered speech response can be provided in addition to performing the task. For example, user **830** may whisper, "Set the reminder for the meeting with John on Tuesday 2 p.m." In response, digital assistant **800** can create a reminder in user **830**'s calendar, and generate and provide a whispered speech response such as "OK, the reminder is set." The generating and providing of a whispered speech response is described in more detail below.

[0255] As illustrated in FIG. **8**A, in some embodiments, whispered speech determination module **820** can determine whether providing a whispered speech response is disabled. For example, user **830** may not wish to receive a whispered speech response in any circumstances and thus may disable the whispered speech response capability of digital assistant **800**. In some examples, in accordance with the determination that providing the whispered speech response is disabled, digital assistant **800** can generate a non-whispered speech response (e.g., a speech response in a regular volume or voice), and provide the non-whispered speech response to user **830**. In other examples, whispered speech determination module **820** can be selectively disabled or enabled in certain circumstances; for example, a user could desire that a whispered speech response always be provided in a particular location or time period.

[0256] As shown in FIG. **8**A, digital assistant **800** can generate a whispered speech response. In some embodiments, digital assistant **800** can generate an intermediate speech based on the speech input, and generate the whispered speech response using the intermediate speech. For example, digital assistant **800** can include a speech synthesis module **840**. As described, speech synthesis module **840** can be implemented in a manner similar to that of speech synthesis module **740** shown in FIG. **7**B. In addition to speech synthesis module **840**, digital assistant **800** can include one or more modules (not shown) such as a STT processing module, a natural language processing module, a dialogue flow processing module, a task flow processing module, and a service processing module. Digital assistant **800** can also include data and models implemented in a manner similar to those included in digital assistant module **726**. As a result, digital assistant **800** can generate text based on the speech input, perform natural language processing of the text, identify a user intent based on a result of the natural language processing, and generate the intermediate speech according to the user intent. For example, digital assistant **800** 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; executing the task flow to fulfill the inferred intent; and generating a response to fulfill the user's intent. The generated response can be sent to speech synthesis module **840** (e.g., speech synthesizer), where it can be processed to generate an intermediate speech.

[0257] In some embodiments, digital assistant **800** can also include a whisper conversion module **810**. Whisper conversion module **810** can generate a whispered speech response using an intermediate speech provided by speech synthesis module **840**. In some examples, the whispered speech response can have substantially the same content as the whispered speech response. For example, both the whispered speech response and the intermediate speech may have the same speech content, such as "The time in Paris right now is 2 p.m.." Unlike the whispered speech response, the intermediate speech may be a non-whispered speech, and as described above, may be different from the whispered speech response in various aspects (e.g., different spectrum characteristics).

[0258] As shown in FIG. **8**A, in some embodiments, whisper conversion module **810** can generate a whispered speech response in accordance with a determination that the whispered speech response is to be provided. For example, upon determining that a whispered speech response is to be provided, whispered speech determination module can transmit signal **822** to whisper conversion module **810**. Upon receiving signal **822**, whisper conversion module **810** can generate a whispered speech response using the intermediate speech.

[0259] FIG. **8**D illustrates a block diagram of the functions of a whisper conversion module **810** according to various examples. In some embodiments, whisper conversion module **810** can include a linear prediction analysis module **812**, a spectrum modification module **814**, and a linear prediction synthesis module **816**.

[0260] As shown in FIG. **8**D, linear prediction analysis module **812** can receive an intermediate speech and generate a residual signal based on a linear prediction analysis of the intermediate speech. In some embodiments, a linear prediction analysis of a speech can estimate one or more formants of the speech, remove the effects of the formants from the speech, and estimate the amplitude and frequency of the remaining signal. The remaining signal can also be referred to as the residual signal. A formant can represent a resonance or a spectral maximum associated with a speech spectrum.

[0261] In some examples, to generate a residual signal, linear prediction analysis module **812** can obtain a plurality of speech frames using the intermediate speech. For example, linear prediction analysis module **812** can perform framing of the intermediate speech such that the intermediate speech is divided into a plurality of speech frames. Each frame can have a corresponding frame length (e.g., 20 ms). As described, the intermediate speech can represent a response to an informational request or to a request to perform a task. The intermediate speech may include a speech signal with varying amplitude. As a result of framing, within each frame, the intermediate speech may be viewed as stationary.

[0262] As illustrated in FIG. **8**D, upon obtaining the plurality of frames, linear prediction analysis module **812** can perform a linear prediction analysis of the speech frames. In some examples, a linear prediction analysis can include pre-emphasis of the speech frames. In processing speech signals, pre-emphasis refers to a process that increases, for example, within a frequency range, the signal magnitude of some frequencies with respect to the signal magnitude of other frequencies. Pre-emphasis can improve the overall signal-to-noise ratio by reducing the adverse effects of phenomena such as attenuation, distortion, or saturation in subsequent signal processing steps. For example, to reduce the adverse effect on the signal-to-noise ratio, linear prediction analysis module **812** can increase the signal magnitude of some frequencies (e.g., higher frequencies) with respect to the signal magnitude of other frequencies (e.g., lower frequencies) of the speech frames.

[0263] In some embodiments, a linear prediction analysis of the speech frames can include estimating a plurality of linear prediction coefficients (LPCs). In some examples, LPCs can be obtained based on spectral envelope information of the speech frames. As a result, LPCs can indicate formants (e.g., a resonance or a spectral maximum) in the speech spectrum and allow the effects of the formants to be removed. LPCs can be represented in line spectral frequencies (LSFs), reflection coefficients (RCs), autocorrelations (ACs), log area ratios (LARs), arcsine of reflection coefficients (ASRCs), impulse responses of LP synthesis filter (IRs), or the like. As described in more detail below, some representations of the LPCs (e.g., LSFs) can be more easily modified. Modified LSFs can be converted to modified LPCs, and used in linear prediction synthesis.

[0264] In some embodiments, estimating the LPCs can include performing a windowing on the pre-emphasized speech frames. A window function may have zero amplitude or reduced amplitude outside of a pre-determined frequency interval. For example, a window function that has constant amplitude inside a frequency interval and zero amplitude outside of the frequency interval is a rectangular window. A window function can include, for example, a rectangular window, a triangular window, a Welch window, a Parzen window, a Bartlett window, a Blackman window, a Dolph-Chebyshev window, a Hamming window, a Hanning (Hann) window, a Kaiser window, or a combination thereof. It is appreciated that in estimating the LPCs, any other window function can be used, and window functions are not limited to the examples provided above.

[0265] As shown in FIG. **8**D, after estimating the LPCs, in some embodiments, a linear prediction analysis includes inverse filtering the pre-emphasized speech frames to obtain a residual signal. In some examples, the inverse filtering is performed based on the LPCs. As described, LPCs can represent the spectral envelope of the speech frames, and indicate formants in the speech frames. As a result, inverse filtering can remove desired formants in the speech frames based on the LPCs. After the formants are removed, the remaining signal is the residual signal. In some embodiments, the residual signal can be further processed, for example, modified and synthesized to generate a whispered speech response, as described in more detail below.

[0266] As illustrated in FIG. **8**D, whisper conversion module **810** can include a spectrum modification module **814**. Spectrum modification module **814** can modify the residual signal to generate a modified residual signal. In some embodiments, spectrum modification module **814** can receive a white noise sequence and the residual signal, estimate energy of the white noise sequence and energy of the residual signal, correlate the energy of the white noise sequence and energy of the residual signal, and compensate the correlated white noise sequence.

[0267] As shown in FIG. **8**D, in some examples, spectrum modification module **814** can receive a white noise sequence. The white noise sequence may be provided by, for example, a white noise generator. A white noise sequence represents a white noise signal and may have a plurality of

frames. A white noise signal is a random signal with a constant power spectral density. A white noise sequence can include a sequence of serially uncorrelated random variables with zero mean and finite variance. In some examples, spectrum modification module **814** can also receive the residual signal. As described, the residual signal can be generated by linear prediction analysis module **812**.

[0268] In response to receiving the white noise sequence and the residual signal, spectrum modification module **814** can estimate the energy of the white noise sequence and the energy of the residual signal. In some embodiments, the residual signal includes a plurality of frames and the white noise sequence also includes a plurality of frames. Spectrum modification module **814** can estimate the energy of the residual frames and the energy of the white noise frames. As shown in FIG. **8**D, based on the estimation, spectrum modification module **814** can correlate the energy of the white noise sequence and the energy of the residual signal. For example, the energy of the plurality of residual frames can be matched to the energy of the corresponding white noise frames. Based on the matching, spectrum modification module **814** can generate a correlated white noise sequence.

[0269] In some embodiments, spectrum modification module **814** can provide the correlated white noise sequence as the modified residual signal to linear prediction synthesis module **816** for speech synthesis. In some embodiments, spectrum modification module **814** can optionally compensate the correlated white noise sequence before providing it for synthesis, such that the synthesized speech response resembles more closely a whispered speech response. In some examples, compensating the correlated white noise sequence can include differentiating, high-pass filter, band-pass filtering, or a combination thereof. For example, the correlated white noise sequence may have a higher magnitude at low frequencies (e.g., less than 800 Hz) than that of a whispered speech response. One or both of differentiating and high-pass filtering can thus be applied to the correlated white noise sequence to compensate this magnitude difference at low frequencies. As another example, the correlated white noise sequence may have different magnitude within a certain mid-frequency range, and band-pass filtering can be applied to compensate such magnitude difference. After performing compensation, spectrum modification module **814** can provide a modified residual signal to linear prediction synthesis module **816**. The modified residual signal may include, for example, the compensated and correlated white noise sequence.

[0270] As illustrated in FIG. **8**D, linear prediction synthesis module **816** can obtain the whispered speech response based on the linear prediction synthesis of the modified residual signal. A linear prediction synthesis can include a process of recombining a plurality of frames (e.g., modified residual frames) to generate a coherent speech signal. In some embodiments, performing the linear prediction synthesis includes generating a plurality of whispered frames using a synthesis filter and the modified LPCs; as well as generating a whispered speech response using the whispered frames.

[0271] In some examples, generating the whispered frames can include processing the modified residual frames with a synthesis filter. Linear prediction synthesis module **816** can generate the synthesis filter. For example, linear prediction synthesis module **816** can obtain the LPCs and modify the LPCs to generate the synthesis filter. A synthesis filter generated using modified LPCs may facilitate the generation of a speech signal that more closely resembles a whispered speech response.

[0272] In some embodiments, linear prediction synthesis module **816** can obtain the LPCs. For example, the LPCs can be provided by linear prediction analysis module **812**. In response to receiving the LPCs, linear prediction synthesis module **816** can modify the LPCs. As described above, LPCs can be represented by LSFs, RCs, ACs, LARs, ASRCs, IRs, or the like. As an example, linear prediction synthesis module **816** can convert LPCs to LSFs. And LSFs can be more easily modified than LPCs. In some embodiments, to resemble more closely the formants in the spectrum of a whispered speech response, linear prediction synthesis module **816** can modify the LSFs such that in the synthesized speech signal, a number of formants in the low-frequency range (e.g., less than 3-4 kHz) are shifted up to a higher-frequency range; a number of formants (other than a first formant) are sharpened in shape; and/or the first formant is shaped to be less sharped. After linear prediction synthesis module **816** modifies the LSFs, it can convert the modified LSFs to modified LPCs, and provide the modified LPCs to generate the synthesis filter.

[0273] In some embodiments, linear prediction synthesis module **816** can generate a plurality of whispered frames. For example, the modified LPCs can be used to generate a synthesis filter. A synthesis filter can be, for example, an interpolation filter associated with upsampling of the modified residual frames associated with the modified residual signal. The synthesis filter can process the modified residual frames to generate the whispered frames.

[0274] In some embodiments, upon generating the whispered frames, linear prediction synthesis module **816** can generate a whispered speech response using the whispered frames. For example, linear prediction synthesis module **816** can perform deframing to combine the plurality of whispered frames to a coherent speech signal. In some examples, deframing can include performing a windowing. As described above, during the linear prediction analysis to generate the residual signal, in some examples, linear prediction analysis module **812** performs a windowing on the pre-emphasized speech frames. The performance of the windowing during the linear prediction analysis can be reversed with the performance of the deframing. For example, during deframing, the whispered frames can be multiplied by a window that is reciprocal to the window used in the earlier linear prediction analysis.

[0275] In some examples, in generating the whispered speech response, linear prediction synthesis module **816** can also perform an overlap-add. An overlap-add enables a more efficient synthesis process such that the whispered speech response can be generated without extra delay (e.g., delay caused by convolution of a very long sequence or signal with a synthesis filter). In some examples, the performance of the overlap-add is optional.

5. Process for Providing a Whispered Speech Response

[0276] FIGS. **9**A-F illustrate a flow diagram of an exemplary process **900** for operating a digital assistant in accordance with some embodiments. Process **900** may be performed using one or more devices **104**, **108**, **200**, **400**, **600**, or **800** (FIGS. **1**, **2**A, **4**, **6**A-B, or **8**A). Operations in process **900** are, optionally, combined or split and/or the order of some operations is, optionally, changed.

[0277] With reference to FIG. **9**A, at block **902**, a speech input can be received from a user. At block **904**, the speech input includes an informational request. At block **906**, the speech input includes a request to perform a task.

[0278] At block **908**, it is optionally determined whether providing a whispered speech response is disabled. At block **910**, in accordance with a determination that providing the whispered speech response is disabled, a non-whispered speech response is generated and provided to the user in lieu of a whispered speech response.

[0279] At block **912**, it is determined, based on the speech input, that a whispered speech response is to be provided. At block **913**, the whispered speech response includes a response to the informational request. At block **914**, the whispered speech response includes a response associated with performing the task.

[0280] At block **916**, in some examples, determining that the whispered speech response is to be provided includes determining whether the speech input includes a whispered speech input. For example, at block **918**, the process may determine that the whispered speech input is associated with a first spectrum having one or more first spectrum characteristics associated with a whispered speech.

[0281] At block **919**, the one or more first spectrum characteristics may include a first amplitude, where, for example, the first amplitude is less than a second amplitude below a threshold frequency, and the second amplitude is associated with the non-whispered speech.

[0282] At block **920**, the one or more first spectrum characteristics include a first energy, where, for example, the first energy is less than a second energy below the threshold frequency, and the second energy is associated with the non-whispered speech.

[0283] At block **921**, the one or more first spectrum characteristics include a first volume, where, for example, the first volume is less than a second volume by a threshold volume percentage, and the second volume is associated with the non-whispered speech.

[0284] At block **922**, the one or more first spectrum characteristics include a first slope of the first spectrum, where, for example, the first slope of the first spectrum is shifted by a threshold slope percentage with respect to a second slope of the second spectrum, and the second slope of the second spectrum is associated with the non-whispered speech.

[0285] With reference to FIG. **9**B, in some examples, at block **924**, it is determined whether the speech input includes a whispered speech input using one or more features of the speech input (e.g., MFCC**0** or MFCC**1**). The one or more features represent one or more spectrum characteristics (e.g., amplitude, energy, volume, slope) associated with a spectrum of the speech input. For example, at block **926**, the spectrum of the speech input is obtained. At block **928**, the one or more spectrum characteristics associated with the spectrum of the speech input are determined. At block **930**, a first feature and a second feature are determined based on the one or more spectrum characteristics associated with the spectrum of the speech input. At block **931**, the first feature is a first mel-frequency cepstrum coefficient (e.g., MFCC**0**) representing an energy or an amplitude associated with the spectrum of the speech input. At block **932**, the second feature is a second mel-frequency cepstrum coefficient (e.g., MFCC**1**) representing the slope associated with the spectrum of the speech input.

[0286] At block **934**, a whisper score of the first feature to the second feature is obtained. At block **936**, it is determined whether the whisper score satisfies a score threshold. At block **937**, the score threshold is adjustable in accordance with the manner in which the user speaks.

[0287] With reference to FIG. **9**C, in some examples, at block **938**, determining that the whispered speech response is to be provided includes determining whether context data indicates that the whispered speech response is expected. At block **939**, the context data is obtained. The context data is provided by at least one of the user device or one or more additional devices communicatively connected to the user device. Upon obtaining the context data, at block **940**, it is determined whether the context data satisfies one or more conditions for providing the whispered speech response.

[0288] At block **942**, in some embodiments, upon determining that a whispered speech response is to be provided, the whispered speech response is generated. At block **944**, generating the whispered speech response includes generating an intermediate speech based on the speech input. At block **945**, the intermediate speech has substantially the same content as the whispered speech response. At block **947**, generating the intermediate speech includes generating text based on the speech input. At block **948**, natural language processing of the text is performed. At block **949**, a user intent is identified based on a result of the natural language processing. At block **950**, the intermediate speech is generated according to the user intent.

[0289] With reference to FIG. **9**D, in some examples, at block **952**, the whispered speech response is generated using the intermediate speech. At block **954**, a residual signal is obtained based on a linear prediction analysis of the intermediate speech. At block **955**, obtaining the residual signal includes obtaining a plurality of speech frames using the intermediate speech. At block **956**, a linear prediction analysis is performed on the plurality of speech frames. At block **958**, performing the linear prediction analysis includes pre-emphasizing the plurality of speech frames. At block **960**, a plurality of LPCs are estimated. At block **962**, estimating the LPCs includes performing a windowing on the pre-emphasized plurality of speech frames. At block **964**, the pre-emphasized speech frames are inverse filtered to obtain the residual signal.

[0290] With reference to FIG. **9**E, at block **966**, in some examples, the residual signal is modified. At block **968**, a white noise sequence is received. At block **969**, the energy of the white noise sequence and the energy of the residual signal are estimated. At block **970**, the energy of the white noise sequence is correlated to the energy of the residual signal. At block **971**, in some examples, the correlated white noise sequence is compensated. At block **972**, compensating the correlated white noise sequence includes performing at least one of differentiating, high-pass filtering, or band-pass filtering with respect to the correlated white noise sequence.

[0291] With reference to FIG. **9**F, at block **974**, in some examples, the whispered speech response is obtained based on the linear prediction synthesis of the modified residual signal. At block **976**, a plurality of LPCs are obtained. At block **978**, the LPCs are modified. At block **979**, modifying the LPCs includes converting the LPCs to line spectral frequencies. At block **980**, the line spectral frequencies are modified. At block **981**, modified LPCs are generated based on the modified line spectral frequencies.

[0292] At block **984**, a linear prediction synthesis is performed using the modified LPCs. At block **985**, a plurality of whispered frames are generated using a synthesis filter and the modified LPCs. At block **986**, the whispered speech response is generated using the plurality of whispered frames. At block **987**, in some examples, generating the whispered speech response includes performing at least one of a windowing or an overlap-add with respect to the plurality of whispered frames.

[0293] FIG. **10** illustrates a flow diagram of an exemplary process **1000** for operating a digital assistant in accordance with some embodiments. Process **1000** may be performed using one or more devices **104**, **108**, **200**, **400**, **600**, or **800** (FIGS. **1**, **2**A, **4**, **6**A-B, or **8**A). Operations in process **1000** are, optionally, combined or split and/or the order of some operations is, optionally, changed.

[0294] With reference to FIG. **10**, at block **1002**, it is detected that the digital assistant is activated. For example, it is detected that the digital assistant is in listening mode.

[0295] At block **1004**, it is determined, based on context data, that a whispered speech response is to be provided. For example, as described above with respect to FIG. **8**A, context data (also referred to as contextual information) can include user-specific data, vocabulary, and/or preferences relevant to the user input. For example, the context data can include the user's calendar, the current time, the user's location data, and/or the user's preferences of operation of a digital assistant (e.g., digital assistant **800**). For example, the user's current location data may indicate that the user is currently physically located inside a library. In some examples, the context data 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 embodiments, determining that a whispered speech response is to be provided is based on context data that is a non-speech input.

[0296] At block **1006**, a speech input is received from the user.

[0297] At block **1008**, upon determining that a whispered speech response is to be provided, a whispered speech response is generated. The generating of the whispered speech response can include those described above with respect to FIGS. **8**A-F. For example, linear prediction analysis and synthesis may be used to generate the whispered speech response.

[0298] At block **1010**, the whispered speech response is provided to the user.

6. Electronic Device

[0299] FIG. **11** shows a functional block diagram of an electronic device **1100** configured in accordance with the principles of the various described examples, including those described with reference to FIGS. **8**A-D, **9**A-F, and **10**. 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. **11** 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.

[0300] As shown in FIG. **11**, electronic device **1100** can include a microphone **1102** and processing unit **1108**. In some examples, processing unit **1108** includes a receiving unit **1110**, an identifying unit **1112**, a generating unit **1114**, and a providing unit **1116**, and optionally, a performing unit **1118**, an identifying unit **1120**, an obtaining unit **1122**, a modifying unit **1124**, a pre-emphasizing unit **1126**, a converting unit **1128**, a compensating unit **1130**, a correlating unit **1132**, an estimating unit **1134**, and an inverse filtering unit **1136**.

[0301] The processing unit **1108** is configured to receive (e.g., with the receiving unit **1110**) a speech input from a user, and determine (e.g., with the determining unit **1112**), based on the speech input, that a whispered speech response is to be provided. The processing unit **1108** is further configured to, upon determining that a whispered speech response is to be provided, generate (e.g., with the generating unit **1114**) the whispered speech response. The processing unit **1108** is further configured to provide (e.g., with the providing unit) the whispered speech response to the user.

[0302] In some examples, the speech input includes an informational request.

[0303] In some examples, the whispered speech response includes a response to the informational request.

[0304] In some examples, the speech input includes a request to perform a task.

[0305] In some examples, the whispered speech response includes a response associated with performing the task.

[0306] In some examples, the determining that the whispered speech response is to be provided includes at least one of: determining (e.g., with the determining unit **1112**) whether the speech input includes a whispered speech input; and determining (e.g., with the determining unit **1112**) whether context data indicates that the whispered speech response is expected.

[0307] In some examples, the whispered speech input is associated with a first spectrum having one or more first spectrum characteristics associated with a whispered speech.

[0308] In some examples, the one or more first spectrum characteristics include a first amplitude. The first amplitude is less than a second amplitude below a threshold frequency (e.g., ~3-4 kHz). The second amplitude is associated with the non-whispered speech.

[0309] In some examples, the one or more first spectrum characteristics include a first energy. The first energy is less than a second energy below the threshold frequency (e.g., 3-4 kHz). The second energy is associated with the non-whispered speech.

[0310] In some examples, the one or more first spectrum characteristics include a first volume. The first volume is less than a second volume by a threshold volume percentage. The second volume is associated with the non-whispered speech.

[0311] In some examples, the one or more first spectrum characteristics include a first slope of the first spectrum. The first slope of the first spectrum is shifted by a threshold slope percentage with respect to a second slope of the second spectrum. The second slope of the second spectrum is associated with the non-whispered speech.

[0312] In some examples, determining whether the speech input includes a whispered speech input includes determining (e.g., with the determining unit **1112**) whether the speech input includes a whispered speech input using one or more

features of the speech input (e.g., MFCC**0** or MFCC**1**). The one or more features represent one or more spectrum characteristics (e.g., amplitude, energy, volume, slope) associated with a spectrum of the speech input.

[0313] In some examples, determining whether the speech input includes a whispered speech input using the one or more features includes obtaining (e.g., with the obtaining unit **1122**) the spectrum of the speech input; determining (e.g., with the determining unit **1112**) the one or more spectrum characteristics associated with the spectrum of the speech input; and determining (e.g., with the determining unit **1112**) a first feature (e.g., MFCC**0**) and a second feature (e.g., MFCC**1**) based on the one or more spectrum characteristics associated with the spectrum of the speech input.

[0314] In some examples, the first feature is a first mel-frequency cepstrum coefficient (e.g., MFCC**0**) representing an energy or an amplitude associated with the spectrum of the speech input. In some examples, the second feature is a second mel-frequency cep strum coefficient (e.g., MFCC**1**) representing the slope associated with the spectrum of the speech input.

[0315] In some examples, processing unit **1308** is further configured to obtain (e.g., with the obtaining unit **1122**) a whisper score based on the first feature to the second feature, and determine (e.g., with the determining unit **1112**) whether the whisper score satisfies a score threshold.

[0316] In some examples, the score threshold is adjustable in accordance with the manner in which the user speaks.

[0317] In some examples, determining whether the context data indicates that the whispered speech response is expected includes obtaining (e.g., with the obtaining unit **1122**) the context data provided by the at least one of the user device or one or more additional devices communicatively connected to the user device; and determining (e.g., with the determining unit **1112**) whether the context data satisfies one or more conditions for providing the whispered speech response.

[0318] In some examples, the processing unit **1108** is further configured to, prior to determining that a whispered speech response is to be provided, determine (e.g., with the determining unit **1112**) whether providing a whispered speech response is disabled (e.g., the user may have a speech disability and may not wish to use this function). In accordance with the determination that providing the whispered speech response is disabled, the processing unit **1108** is further configured to generate (e.g., with the generating unit **1114**) a non-whispered speech response, and provide (e.g., with the providing unit **1116**) the non-whispered speech response to the user in lieu of the whispered speech response.

[0319] In some examples, generating the whispered speech response includes generating (e.g., with the generating unit **1114**) an intermediate speech based on the speech input, and generating (e.g., with the generating unit **1114**) the whispered speech response using the intermediate speech.

[0320] In some examples, the intermediate speech has substantially the same content as the whispered speech response.

[0321] In some examples, generating the intermediate speech based on the speech input includes generating (e.g., with the generating unit **1114**) text based on the speech input; performing (e.g., with the performing unit **1118**) natural language processing of the text; identifying (e.g., with the identifying unit **1120**) a user intent based on a result of the natural language processing; and generating (e.g., with the generating unit **1114**) the intermediate speech according to the user intent.

[0322] In some examples, generating the whispered speech response using the intermediate speech includes obtaining (e.g., with the obtaining unit **1122**) a residual signal based on a linear prediction analysis of the intermediate speech; modifying (e.g., with the modifying unit **1124**) the residual signal; and obtaining (e.g., with the obtaining unit **1122**) the whispered speech response based on a linear prediction synthesis of the modified residual signal.

[0323] In some examples, obtaining the residual signal based on a linear prediction analysis of the intermediate speech includes obtaining (e.g., with the obtaining unit **1122**) a plurality of speech frames using the intermediate speech, and performing (e.g., with the performing unit **1118**) the linear prediction analysis of the plurality of speech frames.

[0324] In some examples, performing the linear prediction analysis of the plurality of speech frames includes pre-emphasizing (e.g., with the pre-emphasizing unit **1126**) the plurality of speech frames; estimating (e.g., with the estimating unit **1134**) a plurality of LPCs; and inverse filtering (e.g., with the inverse filtering unit **1136**) the pre-emphasized speech frames to obtain the residual signal.

[0325] In some examples, estimating the plurality of LPCs includes performing (e.g., with the performing unit **1118**) a windowing on the pre-emphasized plurality of speech frames.

[0326] In some examples, modifying the residual signal includes receiving (e.g., with the receiving unit **1110**) a white noise sequence; estimating (e.g., with the estimating unit **1134**) energy of the white noise sequence and the residual signal; correlating (e.g., with the correlating unit **1132**) the energy of the white noise sequence and the energy of the residual signal; and compensating (e.g., with the compensating unit **1130**) the correlated white noise sequence.

[0327] In some examples, compensating the correlated white noise sequence includes performing (e.g., with the performing unit **1118**) at least one of differentiating, high-pass filtering, or band-pass filtering with respect to the correlated white noise sequence.

[0328] In some examples, obtaining the whispered speech response based on the linear prediction synthesis of the modified residual signal includes obtaining (e.g., with the obtaining unit **1122**) a plurality of LPCs; modifying (e.g., with the modifying unit **1124**) the LPCs; and performing (e.g., with the performing unit **1118**) a linear prediction synthesis of the modified residual signal using the modified LPCs.

[0329] In some examples, performing a linear prediction synthesis of the modified residual signal using the modified LPCs includes generating (e.g., with the generating unit **1114**) a plurality of whispered frames using a synthesis filter and the modified LPCs, and generating (e.g., with the generating unit **1114**) the whispered speech response using the plurality of whispered frames.

[0330] In some examples, generating the whispered speech response includes performing (e.g., with the performing unit **1118**) at least one of a windowing or an overlap-add with respect to the plurality of whispered frames.

[0331] In some examples, the processing unit **1108** can be configured to detect (e.g., with the detecting unit **1138**) that the digital assistant is activated; determine (e.g., with the determining unit **1112**), based on context data, that a whispered speech response is to be provided; receive (e.g., with the receiving unit **1110**) a speech input from a user; upon determining that a whispered speech response is to be provided, generate (e.g., with the generating unit **1114**) a whispered speech response; and provide (e.g., with the providing unit **1116**) the whispered speech response to the user.

[0332] The operation described above with respect to FIG. **11** is, optionally, implemented by the components depicted in FIGS. **1**, **2**A, **4**, **6**A-B, **7**A-**7**B, or **8**A-**8**D. For example, receiving operation **1110**, determining operation **1112**, generating operation **1114**, and providing operation **1116** 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**, **2**A, **4**, **6**A-B, **7**A-**7**B, or **8**A-**8**D.

[0333] It is understood by persons of skill in the art that the functional blocks described in FIG. **11** 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 **1108** can have an associated "controller" unit that is operatively coupled with processing unit **1108** to enable operation. This controller unit is not separately illustrated in FIG. **11** 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 **1108**, such as device **1100**. As another example, one or more units, such as the receiving unit **1110**, may be hardware units outside of processing unit **1108** in some embodiments. The description herein thus optionally supports combination, separation, and/or further definition of the functional blocks described herein.

[0334] For purpose of explanation, the foregoing description 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.

[0335] 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.

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 speech input from a user;
determining whether providing a whispered speech response is disabled;
in accordance with a determination that providing the whispered speech response is not disabled,
determining, based on the speech input, that a whispered speech response is to be provided;
upon determining that a whispered speech response is to be provided, generating the whispered speech response;
providing the whispered speech response to the user;
in accordance with the determination that providing the whispered speech response is disabled,
generating a non-whispered speech response, and
providing the non-whispered speech response to the user in lieu of the whispered speech response.

2. The electronic device of claim **1**, wherein the speech input comprises at least one of an informational request or a request to perform a task.

3. The electronic device of claim **2**, wherein the whispered speech response comprises at least one of a response to the informational request or a response associated with performing the task.

4. The electronic device of claim **1**, wherein determining that the whispered speech response is to be provided comprises at least one of:
determining whether the speech input includes a whispered speech input; and
determining whether context data indicates that the whispered speech response is expected.

5. The electronic device of claim **4**, wherein the whispered speech input is associated with a first spectrum having one or more first spectrum characteristics associated with a whispered speech.

6. The electronic device of claim **5**, wherein the one or more first spectrum characteristics comprise at least one of:
a first amplitude, wherein the first amplitude is less than a second amplitude below a threshold frequency, the second amplitude being associated with the non-whispered speech;
a first energy, wherein the first energy is less than a second energy below the threshold frequency, the second energy being associated with the non-whispered speech;
a first volume, wherein the first volume is less than a second volume by a threshold volume percentage, the second volume being associated with the non-whispered speech; and
a first slope of the first spectrum, wherein the first slope of the first spectrum is shifted by a threshold slope percentage with respect to a second slope of the second spectrum, the second slope of the second spectrum being associated with the non-whispered speech.

7. The electronic device of claim **5**, wherein determining whether the speech input includes a whispered speech input comprises:
determining whether the speech input includes a whispered speech input using one or more features of the speech input, wherein the one or more features represent one or more spectrum characteristics associated with a spectrum of the speech input.

8. The electronic device of claim **7**, wherein determining whether the speech input includes a whispered speech input using the one or more features comprises:
obtaining the spectrum of the speech input;
determining the one or more spectrum characteristics associated with the spectrum of the speech input; and

determining a first feature and a second feature based on the one or more spectrum characteristics associated with the spectrum of the speech input.

9. The electronic device of claim **8**,

wherein the first feature is a first mel-frequency cepstrum coefficient (MFCC**0**) representing an energy or an amplitude associated with the spectrum of the speech input; and

wherein the second feature is a second mel-frequency cepstrum coefficient (MFCC**1**) representing the slope associated with the spectrum of the speech input.

10. The electronic device of claim **8**, further comprising:

obtaining a whisper score based on the first feature to the second feature; and

determining whether the whisper score satisfies a score threshold.

11. The electronic device of claim **4**, wherein determining whether the context data indicates that the whispered speech response is expected comprises:

obtaining the context data provided by the at least one of the user device or one or more additional devices communicatively connected to the user device; and

determining whether the context data satisfy one or more conditions for providing the whispered speech response.

12. (canceled)

13. The electronic device of claim **1**, wherein generating the whispered speech response comprises:

generating an intermediate speech based on the speech input; and

generating the whispered speech response using the intermediate speech.

14. The electronic device of claim **13**, wherein the intermediate speech has substantially the same content as the whispered speech response.

15. The electronic device of claim **13**, wherein generating the intermediate speech based on the speech input comprises:

generating text based on the speech input;

performing natural language processing of the text;

identifying a user intent based on a result of the natural language processing; and

generating the intermediate speech according to the user intent.

16. The electronic device of claim **13**, wherein generating the whispered speech response using the intermediate speech comprises:

obtaining a residual signal based on a linear prediction analysis of the intermediate speech;

modifying the residual signal; and

obtaining the whispered speech response based on a linear prediction synthesis of the modified residual signal.

17. The electronic device of claim **16**, wherein obtaining the residual signal based on a linear prediction analysis of the intermediate speech comprises:

obtaining a plurality of speech frames using the intermediate speech; and

performing the linear prediction analysis of the plurality of speech frames.

18. The electronic device of claim **17**, wherein performing the linear prediction analysis of the plurality of speech frames comprises:

pre-emphasizing the plurality of speech frames;

estimating a plurality of linear prediction coefficients; and

inverse filtering the pre-emphasized speech frames to obtain the residual signal.

19. The electronic device of claim **18**, wherein estimating the plurality of linear prediction coefficients comprises:

performing a windowing on the pre-emphasized plurality of speech frames.

20. The electronic device of claim **16**, wherein modifying the residual signal comprises:

receiving a white noise sequence;

estimating energy of the white noise sequence and the residual signal;

correlating the energy of the white noise sequence and the energy of the residual signal; and

compensating the correlated white noise sequence.

21. The electronic device of claim **20**, wherein compensating the correlated white noise sequence comprises performing at least one of differentiating, high-pass filtering, or band-pass filtering with respect to the correlated white noise sequence.

22. The electronic device of claim **16**, wherein obtaining the whispered speech response based on the linear prediction synthesis of the modified residual signal comprises:

obtaining a plurality of linear prediction coefficients;

modifying the linear prediction coefficients; and

performing a linear prediction synthesis of the modified residual signal using the modified linear prediction coefficients.

23. The electronic device of claim **22**, wherein modifying the linear prediction coefficients comprises:

converting the plurality of linear prediction coefficients to line spectral frequencies;

modifying the line spectral frequencies; and

generating modified linear prediction coefficients based on the modified line spectral frequencies.

24. The electronic device of claim **22**, wherein performing a linear prediction synthesis of the modified residual signal using the modified linear prediction coefficients comprises:

generating a plurality of whispered frames using a synthesis filter and the modified linear prediction coefficients; and

generating the whispered speech response using the plurality of whispered frames.

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 speech input from a user;

determining whether providing a whispered speech response is disabled;

in accordance with a determination that providing the whispered speech response is not disabled,

determine, based on the speech input, that a whispered speech response is to be provided;

upon determining that a whispered speech response is to be provided, generate the whispered speech response;

provide the whispered speech response to the user;

in accordance with the determination that providing the whispered speech response is disabled,

generating a non-whispered speech response, and

providing the non-whispered speech response to the user in lieu of the whispered speech response.

26. A method for operating a digital assistant, comprising:

at a user device with one or more processors and memory:

receiving a speech input from a user;

in accordance with a determination that providing the whispered speech response is not disabled,

determining, based on the speech input, that a whispered speech response is to be provided;

upon determining that a whispered speech response is to be provided, generating the whispered speech response; and

providing the whispered speech response to the user;

in accordance with the determination that providing the whispered speech response is disabled,

generating a non-whispered speech response, and

providing the non-whispered speech response to the user in lieu of the whispered speech response.

* * * * *