



(51) International Patent Classification:
G10L 15/22 (2006.01)

(21) International Application Number:
PCT/US2015/047553

(22) International Filing Date:
28 August 2015 (28.08.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/057,979 30 September 2014 (30.09.2014) US
14/834,239 24 August 2015 (24.08.2015) US

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

(72) Inventor: **KIM, Yoon;** c/o Apple Inc., 1 Infinite Loop, MS 169-IPL, Cupertino, CA 95014 (US).

(74) Agents: **EIDE, Christopher, B.** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

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

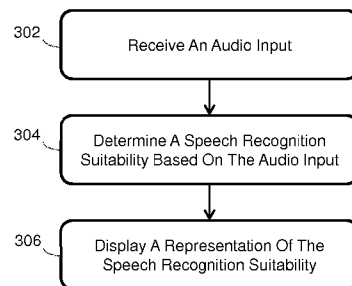
(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: PROVIDING AN INDICATION OF THE SUITABILITY OF SPEECH RECOGNITION

Process
300



(57) Abstract: This relates to providing an indication of the suitability of an acoustic environment for performing speech recognition. One process can include receiving an audio input and determining a speech recognition suitability based on the audio input. The speech recognition suitability can include a numerical, textual, graphical, or other representation of the suitability of an acoustic environment for performing speech recognition. The process can further include displaying a visual representation of the speech recognition suitability to indicate the likelihood that a spoken user input will be interpreted correctly. This allows a user to determine whether to proceed with the performance of a speech recognition process, or to move to a different location having a better acoustic environment before performing the speech recognition process. In some examples, the user device can disable operation of a speech recognition process in response to determining that the speech recognition suitability is below a threshold suitability.

FIG. 3

PROVIDING AN INDICATION OF THE SUITABILITY OF SPEECH RECOGNITION

Cross-Reference to Related Application

[0001] This application claims priority from U.S. Provisional Serial No. 62/057,979, filed on September 30, 2014, entitled PROVIDING AN INDICATION OF THE SUITABILITY OF SPEECH RECOGNITION, and U.S. Non-Provisional Serial No. 14/834,239, filed on August 24, 2015, entitled PROVIDING AN INDICATION OF THE SUITABILITY OF SPEECH RECOGNITION, which are hereby incorporated by reference in their entirety for all purposes.

Field

[0002] This relates generally to natural language processing and, more specifically, to determining the suitability of an acoustic environment for performing speech recognition.

Background

[0003] Intelligent automated assistants (or virtual assistants) provide an intuitive interface between users and electronic devices. These assistants can allow users to interact with devices or systems using natural language in spoken and/or text forms. For example, a user can access the services of an electronic device by providing a spoken user input in natural language form to a virtual assistant associated with the electronic device. The virtual assistant can perform natural language processing on the spoken user input to infer the user's intent and operationalize the user's intent into tasks. The tasks can then be performed by executing one or more functions of the electronic device, and a relevant output can be returned to the user in natural language form.

[0004] The acoustic environment in which the virtual assistant operates can affect the virtual assistant's ability to interpret a user's spoken input. For example, background noise, such as music, conversations of other individuals, traffic noises, or the like, can obscure the user's spoken input contained in the audio received by the virtual assistant. This can result in words being interpreted incorrectly or not at all. Thus, it can be desirable to operate a virtual assistant in an acoustic environment that is conducive to performing speech recognition.

Summary

[0005] Systems and processes for providing an indication of the suitability of an acoustic environment for performing speech recognition are disclosed. One process can include receiving an audio input and determining a speech recognition suitability based on the audio input. The speech recognition suitability can include a numerical, textual, graphical, or other representation of the suitability of an acoustic environment for performing speech recognition. The process can further include displaying a visual representation of the speech recognition suitability to indicate the likelihood that a spoken user input will be interpreted correctly. This allows a user to determine whether to proceed with the performance of a speech recognition process, or to move to a different location having a better acoustic environment before performing the speech recognition process. In some examples, the user device can disable operation of a speech recognition process in response to determining that the speech recognition suitability is below a threshold suitability.

Brief Description of the Drawings

[0006] FIG. 1 illustrates an exemplary environment in which a virtual assistant can operate according to various examples.

[0007] FIG. 2 illustrates an exemplary user device according to various examples.

[0008] FIG. 3 illustrates an exemplary process for providing a visual representation of the suitability of an acoustic environment for performing speech recognition according to various examples.

[0009] FIGs. 4-7 illustrate exemplary visual representations of the suitability of an acoustic environment for performing speech recognition according to various examples.

[0010] FIG. 8 illustrates an exemplary process for operating a virtual assistant based on a determined suitability of an acoustic environment for performing speech recognition according to various examples.

[0011] FIG. 9 illustrates a functional block diagram of an electronic device configured to determine a suitability of an acoustic environment for performing speech recognition and for providing a visual representation of the same according to various examples.

Detailed Description

[0012] In the following description of examples, reference is made to the accompanying drawings in which it is shown by way of illustration specific examples that can be practiced. It is to be understood that other examples can be used and structural changes can be made without departing from the scope of the various examples.

[0013] This relates to providing an indication of the suitability of an acoustic environment for performing speech recognition. One process can include receiving an audio input and determining a speech recognition suitability based on the audio input. The speech recognition suitability can include a numerical, textual, graphical, or other representation of the suitability of an acoustic environment for performing speech recognition. The process can further include displaying a visual representation of the speech recognition suitability to indicate the likelihood that a spoken user input will be interpreted correctly. This allows a user to determine whether to proceed with the performance of a speech recognition process, or to move to a different location having a better acoustic environment before performing the speech recognition process. In some examples, the user device can disable operation of a speech recognition process in response to determining that the speech recognition suitability is below a threshold suitability.

SYSTEM OVERVIEW

[0014] FIG. 1 illustrates exemplary system 100 for implementing a virtual assistant according to various examples. The terms “virtual assistant,” “digital 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.

[0015] A virtual 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 seeks either an informational answer or performance of a task by the virtual assistant. A satisfactory response to the user request can include provision of the

requested informational answer, performance of the requested task, or a combination of the two. For example, a user can ask the virtual assistant a question, such as “Where am I right now?” Based on the user’s current location, the virtual assistant can answer, “You are in Central Park.” The user can also request the performance of a task, for example, “Please remind me to call Mom at 4 p.m. today.” In response, the virtual assistant can acknowledge the request and then create an appropriate reminder item in the user’s electronic schedule. During the performance of a requested task, the virtual 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 virtual assistant to request information or performance of various tasks. In addition to providing verbal responses and taking programmed actions, the virtual assistant can also provide responses in other visual or audio forms (e.g., as text, alerts, music, videos, animations, etc.).

[0016] An example of a virtual assistant is described in Applicants’ U.S. Utility Application Serial No. 12/987,982 for “Intelligent Automated Assistant,” filed January 10, 2011, the entire disclosure of which is incorporated herein by reference.

[0017] As shown in FIG. 1, in some examples, a virtual assistant can be implemented according to a client-server model. The virtual assistant can include a client-side portion executed on a user device 102, and a server-side portion executed on a server system 110. User device 102 can include any electronic device, such as a mobile phone, tablet computer, portable media player, desktop computer, laptop computer, PDA, television, television set-top box, wearable electronic device, or the like, and can communicate with server system 110 through one or more networks 108, which can include the Internet, an intranet, or any other wired or wireless public or private network. The client-side portion executed on user device 102 can provide client-side functionalities, such as user-facing input and output processing and communications with server system 110. Server system 110 can provide server-side functionalities for any number of clients residing on a respective user device 102.

[0018] Server system 110 can include one or more virtual assistant servers 114 that can include a client-facing I/O interface 122, one or more processing modules 118, data and model storage 120, and an I/O interface to external services 116. The client-facing I/O interface 122 can facilitate the client-facing input and output processing for virtual assistant server 114. The one or more processing modules 118 can utilize data and model storage 120 to determine the user’s intent based on natural language input, and perform task execution based on inferred user intent. In some examples, virtual assistant server 114 can

communicate with external services 124, such as telephony services, calendar services, information services, messaging services, navigation services, and the like, through network(s) 108 for task completion or information acquisition. The I/O interface to external services 116 can facilitate such communications.

[0019] Server system 110 can be implemented on one or more standalone data processing devices or a distributed network of computers. In some examples, server system 110 can 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 110.

[0020] Although the functionality of the virtual assistant is shown in FIG. 1 as including both a client-side portion and a server-side portion, in some examples, the functions of the assistant can be implemented as a standalone application installed on a user device. In addition, the division of functionalities between the client and server portions of the virtual assistant can vary in different examples. For instance, in some examples, the client executed on user device 102 can be a thin-client that provides only user-facing input and output processing functions, and delegates all other functionalities of the virtual assistant to a backend server.

USER DEVICE

[0021] FIG. 2 is a block diagram of a user-device 102 according to various examples. As shown, user device 102 can include a memory interface 202, one or more processors 204, and a peripherals interface 206. The various components in user device 102 can be coupled together by one or more communication buses or signal lines. User device 102 can further include various sensors, subsystems, and peripheral devices that are coupled to the peripherals interface 206. The sensors, subsystems, and peripheral devices gather information and/or facilitate various functionalities of user device 102.

[0022] For example, user device 102 can include a motion sensor 210, a light sensor 212, and a proximity sensor 214 coupled to peripherals interface 206 to facilitate orientation, light, and proximity sensing functions. One or more other sensors 216, such as a positioning system (e.g., a GPS receiver), a temperature sensor, a biometric sensor, a gyroscope, a compass, an accelerometer, and the like, are also connected to peripherals interface 206, to facilitate related functionalities.

[0023] In some examples, a camera subsystem 220 and an optical sensor 222 can be utilized to facilitate camera functions, such as taking photographs and recording video clips. Communication functions can be facilitated through one or more wired and/or wireless communication subsystems 224, which can include various communication ports, radio frequency receivers and transmitters, and/or optical (e.g., infrared) receivers and transmitters. An audio subsystem 226 can be coupled to speakers 228 and a microphone 230 to facilitate voice-enabled functions, such as voice recognition, voice replication, digital recording, and telephony functions.

[0024] In some examples, user device 102 can further include an I/O subsystem 240 coupled to peripherals interface 206. I/O subsystem 240 can include a touch screen controller 242 and/or other input controller(s) 244. Touch-screen controller 242 can be coupled to a touch screen 246. Touch screen 246 and the touch screen controller 242 can, for example, detect contact and movement or break thereof using any of a plurality of touch sensitivity technologies, such as capacitive, resistive, infrared, and surface acoustic wave technologies, proximity sensor arrays, and the like. Other input controller(s) 244 can be coupled to other input/control devices 248, such as one or more buttons, rocker switches, a thumb-wheel, an infrared port, a USB port, and/or a pointer device such as a stylus.

[0025] In some examples, user device 102 can further include a memory interface 202 coupled to memory 250. Memory 250 can include any electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, a portable computer diskette (magnetic), a random access memory (RAM) (magnetic), a read-only memory (ROM) (magnetic), an erasable programmable read-only memory (EPROM) (magnetic), a portable optical disc such as CD, CD-R, CD-RW, DVD, DVD-R, or DVD-RW, or flash memory such as compact flash cards, secured digital cards, USB memory devices, memory sticks, and the like. In some examples, a non-transitory computer-readable storage medium of memory 250 can be used to store instructions (e.g., for performing some or all of process 300 or 800, 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 process 300 or 800, described below) can be stored on a non-transitory computer-readable storage medium of server system 110, or can be divided between the non-transitory computer-readable storage medium of memory 250 and the non-transitory

computer-readable storage medium of server system 110. 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.

[0026] In some examples, the memory 250 can store an operating system 252, a communication module 254, a graphical user interface module 256, a sensor processing module 258, a phone module 260, and applications 262. Operating system 252 can include instructions for handling basic system services and for performing hardware dependent tasks. Communication module 254 can facilitate communicating with one or more additional devices, one or more computers, and/or one or more servers. Graphical user interface module 256 can facilitate graphic user interface processing. Sensor processing module 258 can facilitate sensor related processing and functions. Phone module 260 can facilitate phone-related processes and functions. Application module 262 can facilitate various functionalities of user applications, such as electronic-messaging, web browsing, media processing, navigation, imaging, and/or other processes and functions.

[0027] As described herein, memory 250 can also store client-side virtual assistant instructions (e.g., in a virtual assistant client module 264) and various user data 266 (e.g., user-specific vocabulary data, preference data, and/or other data, such as the user’s electronic address book, to-do lists, shopping lists, etc.) to provide the client-side functionalities of the virtual assistant.

[0028] In various examples, virtual assistant client module 264 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., I/O subsystem 240, audio subsystem 226, or the like) of user device 102. Virtual assistant client module 264 can also be capable of providing output in audio (e.g., speech output), visual, and/or tactile forms. 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, virtual assistant client module 264 can communicate with the virtual assistant server using communication subsystem 224.

[0029] In some examples, virtual assistant client module 264 can utilize the various sensors, subsystems, and peripheral devices to gather additional information from the surrounding environment of user device 102 to establish a context associated with a user, the

current user interaction, and/or the current user input. In some examples, virtual assistant client module 264 can provide the contextual information or a subset thereof with the user input to the virtual assistant server to help infer the user's intent. The virtual assistant can also use the contextual information to determine how to prepare and deliver outputs to the user.

[0030] In some examples, the contextual information that accompanies the user input can include sensor information, such as lighting, ambient noise, ambient temperature, images or videos of the surrounding environment, distance to another object, and the like. The contextual information can further include information associated with the physical state of user device 102 (e.g., device orientation, device location, device temperature, power level, speed, acceleration, motion patterns, cellular signal strength, etc.) or the software state of user device 102 (e.g., running processes, installed programs, past and present network activities, background services, error logs, resources usage, etc.). Any of these types of contextual information can be provided to the virtual assistant server 114 as contextual information associated with a user input.

[0031] In some examples, virtual assistant client module 264 can selectively provide information (e.g., user data 266) stored on user device 102 in response to requests from the virtual assistant server 114. Virtual assistant client module 264 can also elicit additional input from the user via a natural language dialogue or other user interfaces upon request by virtual assistant server 114. Virtual assistant client module 264 can pass the additional input to virtual assistant server 114 to help virtual assistant server 114 in intent inference and/or fulfillment of the user's intent expressed in the user request.

[0032] In various examples, memory 250 can include additional instructions or fewer instructions. Furthermore, various functions of user device 102 can be implemented in hardware and/or in firmware, including in one or more signal processing and/or application specific integrated circuits.

DETERMINING ACOUSTIC ENVIRONMENT SUITABILITY

[0033] As mentioned above, the acoustic environment in which a virtual assistant operates can affect the virtual assistant's ability to interpret a user's spoken input. For example, background noise, such as music, conversations of other individuals, traffic noises, or the like, can obscure the user's spoken input contained in the audio received by the virtual assistant. Thus, it can be desirable to operate a virtual assistant in an acoustic environment

that is conducive to performing speech recognition. FIG. 3 illustrates an exemplary process for determining a suitability of an acoustic environment for performing speech recognition and providing a visual representation of the same according to various examples.

[0034] At block 302, an audio input can be received at a user device. In some examples, a user device (e.g., user device 102) can receive the audio input via a microphone (e.g., microphone 230). The microphone can convert the audio input into an analog or digital representation, and provide audio data representing the audio input to one or more processors (e.g., processor(s) 204) of the user device. In some examples, the audio input can include user speech. In other examples, the audio input may not include user speech.

[0035] At block 304, a speech recognition suitability can be determined based on the audio input received at block 302. The speech recognition suitability can include a numerical (e.g., a score from 1-5, 1-100, etc.), textual or categorical (e.g., “poor,” “fair,” “good,” “excellent,” “acceptable,” “unacceptable,” etc.), graphical (e.g., a color image), or other representation of the suitability or conduciveness of an acoustic environment in which the user device is located for performing speech recognition. In other words, the speech recognition suitability can generally indicate whether a speech recognition process performed in a particular acoustic environment is likely to produce accurate or inaccurate results.

[0036] In some examples, determining the speech recognition suitability can generally include determining one or more characteristics of the acoustic environment in which the user device is located and determining the speech recognition suitability based on those one or more characteristics. The characteristics of the acoustic environment can include any desired characteristic, such as a signal-to-noise ratio (SNR), an SNR in one or more frequency bands, an intensity (e.g., volume) and/or type of noise detected (e.g., music, speech, pink noise, transient noise, etc.), an intensity and/or type of noise detected in one or more frequency bands, a number of occurrences and intensity of a particular type of noise detected during a previous length of time, or the like. While specific characteristics of an acoustic environment are provided above, it should be appreciated that any other characteristic of an acoustic environment that is indicative of the likelihood that the acoustic environment will negatively affect the performance of a speech recognition process can be used at block 304 to determine the speech recognition suitability.

[0037] In some examples, when using the SNR or the SNR of one or more frequency bands as one of the characteristics to determine the speech recognition suitability of the

acoustic environment, block 304 can include determining an average noise level and an average signal level over any desired length of time based on the audio input received at block 302. This can be performed over all frequencies or can be performed for multiple discrete frequency bands. These values can then be used to compute an SNR over all frequencies or for the multiple discrete frequency bands for the acoustic environment. In general, a higher SNR can be indicative of an acoustic environment that is more suitable or conducive to performing speech recognition and can cause the speech recognition suitability determined at block 304 to be a value that indicates a more suitable acoustic environment. In contrast, a lower SNR can be indicative of an acoustic environment that is less suitable or conducive to performing speech recognition and can cause the speech recognition suitability determined at block 304 to be a value that indicates a less suitable acoustic environment. Examples of how the SNR can be used to calculate the speech recognition suitability are discussed in greater detail below.

[0038] In some examples, when using the type of noise detected or the type of noise detected in one or more frequency bands as one of the characteristics to determine the speech recognition suitability of the acoustic environment, block 304 can include inputting the audio received at block 302 into one or more audio classifiers to determine a type of a detected noise. For example, audio classifiers can be used to identify music, speech, transient noises (e.g., sounds that last for short duration, such as a honk of a car's horn, the beating of a drum, the sound of an object falling to the ground, etc.), pink noise, or the like. This can be performed over all frequencies or can be performed for multiple discrete frequency bands. The type(s) of identified noises can then be used to compute a speech recognition suitability. In general, the influence of the identified type of noise on the speech recognition suitability can be related to the difficulty of filtering the type of noise or the likelihood that the type of noise will negatively affect the performance of a speech recognition process. For example, non-transient noises that are slow-varying in time (e.g., wind, sound of a car driving on the road, etc.) can often be filtered from the audio input, causing a minimal impact on a speech recognition process. As a result, these types of noises can result in less of a negative impact on the determined speech recognition suitability. However, transient noises can be more difficult to filter from an audio input, causing a greater impact on a speech recognition process. As a result, these types of noises can result in a greater negative impact on the determined speech recognition suitability. In some examples, the identified type of noise can be combined with a detected intensity of the noise. For example, noises with higher

intensities can more greatly impact the performance of a speech recognition process, while noises with lower intensities can have less of an impact. Examples of how the intensity and/or type of noise can be used to calculate the speech recognition suitability are discussed in greater detail below.

[0039] In some examples, when using the number of occurrences and intensity of a particular type of noise detected during a previous length of time as one of the characteristics to determine the speech recognition suitability of the acoustic environment, block 304 can include maintaining a buffer that at least includes the last X minutes of audio input received at block 302 and analyzing the buffer to determine the number of occurrences and volume of a particular type of noise detected within the last X minutes of time. For example, block 304 can include determining the number of occurrences and volume of transient type noises within last X minutes. This value can be indicative of the likelihood that a similar type of noise will occur in the future. In general, a higher intensity and larger number of detected occurrences of the particular type of noise can be indicative of an acoustic environment that is less suitable or conducive to performing speech recognition and can cause the speech recognition suitability determined at block 304 to be a value that indicates a less suitable acoustic environment. In contrast, a lower intensity and smaller number of detected occurrences of the particular type of noise can be indicative of an acoustic environment that is more suitable or conducive to performing speech recognition and can cause the speech recognition suitability determined at block 304 to be a value that indicates a more suitable acoustic environment. Examples of how the number of occurrences and intensity of a particular type of noise detected during a previous length of time can be used to calculate the speech recognition suitability are discussed in greater detail below.

[0040] The speech recognition suitability can be determined using one or more characteristics of the acoustic environment, such as those described above, in various ways. For instance, in some examples, each characteristic of the acoustic environment can be assigned a speech recognition contribution value that is representative of a likelihood that an acoustic environment having that characteristic is to negatively affect the performance of a speech recognition process. In general, a characteristic that is indicative of an acoustic environment that is suitable or conducive to performing speech recognition can be assigned a low speech recognition contribution value (or a negative value), while a characteristic that is indicative of an acoustic environment that is not suitable or conducive to performing speech recognition can be assigned a large, positive speech recognition contribution value. The

speech recognition contribution values can be combined with the speech recognition contribution values associated with other characteristics of the acoustic environment to calculate an overall speech recognition suitability score. For example, the speech recognition contribution values associated with each characteristic of the acoustic environment can be combined by calculating the sum of the values, calculating a weighted sum of the values, or combining the values in any other desired manner to generate an overall speech recognition suitability score. In this example, a large overall speech recognition suitability score can be indicative of an acoustic environment that is not suitable or conducive to performing speech recognition, while a small (or negative) overall speech recognition suitability score can be indicative of an acoustic environment that is suitable or conducive to performing speech recognition. It should be appreciated that, in other examples, different numbering conventions can be used such that a small (or negative) overall speech recognition suitability score can be indicative of an acoustic environment that is not suitable or conducive to performing speech recognition, while a large, positive overall speech recognition suitability score can be indicative of an acoustic environment that is suitable or conducive to performing speech recognition.

[0041] To illustrate, the speech recognition contribution value assigned to an SNR characteristic can be the inverse of the SNR value, can be the inverse of the SNR value multiplied by a scaling factor, or can be a different value assigned using a different convention that generally relates the assigned value to the likelihood that an acoustic environment having the SNR value is suitable or conducive to performing speech recognition. For example, the speech recognition contribution value can vary monotonically (linearly or non-linearly) with respect to the inverse of the SNR value such that smaller SNR values can be assigned larger speech recognition contribution values.

[0042] Similarly, the speech recognition contribution value assigned to a noise-type characteristic can be a value that is assigned to each of the different types of detectable noises. For example, pink noises, which can be easy to filter and can have a low impact on the performance of a speech recognition process, can be assigned lower speech recognition contribution values, while transient noises, which can have significant negative effects on a speech recognition process, can be assigned larger speech recognition contribution values. Other types of noises can similarly be assigned speech recognition contribution values based on the likelihood that those types of noises will negatively impact the performance of a speech recognition process. In some examples, the intensity (e.g., volume) of the noise can

also be used to adjust the speech recognition contribution value assigned to the type of detected noise. For example, a noise having a lower intensity can be used to reduce the negative impact of the noise on the determined speech recognition suitability (e.g., by multiplying the intensity by the value assigned to the type of noise), while a noise having a higher intensity can be used to increase the negative impact of the noise on the determined speech recognition suitability (e.g., by multiplying the intensity by the value assigned to the type of noise).

[0043] Additionally, the speech recognition contribution value assigned to the characteristic indicating the number of occurrences and intensity of a particular type of noise detected during a previous length of time can depend on the number of occurrences and intensity of the detected noises. For example, a larger number of occurrences can be associated with a larger speech recognition contribution value, while a lower number of occurrences can be associated with a smaller speech recognition contribution value. In some examples, the intensity (e.g., volume) of the noise can also be used to adjust the speech recognition contribution value assigned to the number of detected noises. For example, a noise having a lower intensity can be used to reduce the negative impact of the noise on the determined speech recognition suitability (e.g., by multiplying the intensity by the number of occurrences of the type of noise), while a noise having a higher intensity can be used to increase the negative impact of the noise on the determined speech recognition suitability (e.g., by multiplying the intensity by the value assigned to the type of noise).

[0044] The speech recognition contribution values for the various characteristics of the acoustic environment can then be combined as mentioned above (e.g., by calculating the sum of the values, calculating a weighted sum of the values, or combining the values in any other desired manner) to generate an overall speech recognition suitability score.

[0045] In other examples, the speech recognition suitability can be determined by determining a speech recognition suitability category based on the characteristics of the acoustic environment. In some examples, conditional logic based on one or more characteristics of the acoustic environment can be used to select one of multiple speech recognition suitability categories (e.g., “poor,” “fair,” “good,” “excellent,” “acceptable,” “unacceptable,” etc.) for a particular environment. For example, conditions can be established that state if the SNR value is within a first range of values, then the speech recognition suitability can be assigned a first category, that state that if the SNR value is within a second range of values, then the speech recognition suitability can be assigned a

second category, and so on. Similarly, conditions can be established that state if the type of detected noise is a first type of noise (e.g., pink noise), then the speech recognition suitability can be assigned a first category (e.g., “fair”), that state if the type of detected noise is a second type of noise (e.g., transient noise), then the speech recognition suitability can be assigned a second category (e.g., “poor”), and so on. Similarly, conditions can be established that state if the number of occurrences of a particular type of noise (and, optionally, combined with the intensity of the noise) is within a first range of values, then the speech recognition suitability can be assigned a first category, that state that if the number of occurrences of a particular type of noise (and, optionally, combined with the intensity of the noise) is within a second range of values, then the speech recognition suitability can be assigned a second category, and so on. In other examples, conditions can be established that depend on any two or more characteristics of the acoustic environment. For example, one condition can be that if the SNR value is below a lower threshold value and if the number of transient noises detected in the last X minutes is greater than an upper threshold value, then the speech recognition suitability can be assigned a category of “poor.” It should be appreciated that any number of conditions that depend on any number of characteristics can similarly be used to determine the speech recognition suitability.

[0046] In yet other examples, the speech recognition suitability can be generated through the use of a neural network. In these examples, a speech recognition suitability vector can be determined based on the audio input received at block 302. The speech recognition suitability vector can be expressed as: vector = [S1, S2, . . . SN], where each element S represents a different characteristic of the acoustic environment as determined from the audio input received at block 302. For example, a first element S1 can represent an SNR for a first frequency band, a second element S2 can represent an SNR for a second frequency band, a third element S3 can represent a type of noise detected (e.g., over all frequencies or for a particular frequency), a fourth element S4 can represent an intensity of the detected noise, a fifth element S5 can include a number of occurrences of particular type of noise (e.g., transient noises) detected over the last X minutes, a sixth element S6 can represent an intensity of one or more of the occurrences of the particular type of noise, and so on. The vector can include any number of elements and the values assigned to the elements can be assigned using any desired convention, such as the speech recognition contribution values assigned using the first method of determining speech recognition suitability discussed above.

[0047] The speech recognition suitability vector can be input into the neural network, which can then determine an overall speech recognition suitability based on the values of the elements contained in the speech recognition suitability vector. The result can be a numerical representation (e.g., a score from 1-5, 1-100, etc.), a textual or categorical representation (e.g., “poor,” “fair,” “good,” “excellent,” “acceptable,” “unacceptable,” etc.), a graphical representation (e.g., a color), or any other desired representation of the suitability or conduciveness of the acoustic environment for performing speech recognition. In these examples, the neural network can be trained using vectors of acoustic environments that have been determined to be suitable or unsuitable for performing speech recognition to cause the neural network to provide a desired speech recognition suitability for any combination of characteristic values of the acoustic environment.

[0048] In some examples, the characteristics of the acoustic environment can be determined by hardware within the user device (e.g., processors or other circuitry configured to perform the task of determining the characteristics). For example, the hardware can include classifiers to determine a type of noise, circuitry to measure an intensity of noise, and counters to track the number of occurrences of a particular type of noise over time. The hardware can periodically, intermittently, or at any other desired interval of time output tagged values that can be processed using software on the user device to determine the speech recognition suitability. This hardware implementation advantageously reduces that amount of power required to determine the speech recognition suitability over extended periods of time. For example, if the user device is configured to continuously (or intermittently over an extended period of time) monitor an acoustic environment by continuing to receive and monitor an audio input (e.g., at block 302), determining the characteristics of the acoustic environment using hardware specifically designed to perform that function can result in increased power efficiency.

[0049] At block 306, the user device can display a visual representation of the speech recognition suitability. The visual representation can include any desired representation. For example, FIG. 4 illustrates an example interface 400 that can be displayed at block 306 of process 300. As shown, interface 400 includes visual representation 402 that includes multiple vertical bars, where the value of the speech recognition suitability is represented by a number of the one or more bars that are filled-in. For instance, in this example, the speech recognition suitability can be mapped to one of five (or any other desired number) discrete values (e.g., 0, 1, 2, 3, or 4) corresponding to the number of possible bars that can be filled-in.

For example, if the speech recognition suitability values range from 1-10, the speech recognition suitability can be mapped linearly to the five discrete values (e.g., a speech recognition of 1-2 corresponds to 4 bars, a speech recognition of 3-4 corresponds to 3 bars, a speech recognition of 5-6 corresponds to 2 bars, a speech recognition of 7-8 corresponds to 1 bar, and a speech recognition of 9-10 corresponds to 0 bars). In other examples, the speech recognition suitability can be mapped non-linearly to the discrete number of bars.

[0050] In other examples, the visual representation can include text representing the value of the speech recognition suitability. For example, FIG. 5 illustrates interface 500 that can be displayed at block 306 of process 300. As shown, interface 500 includes visual representation 502 in the form of the text “GOOD” that represents the value of the speech recognition suitability. In this example, the speech recognition suitability can be mapped to one of four (or any other desired number) different descriptions (e.g., “poor,” “fair,” “good,” “excellent,”). For example, if the speech recognition suitability values range from 1-20, the speech recognition suitability can be mapped linearly to the four discrete descriptions (e.g., a speech recognition of 1-5 corresponds to “EXCELLENT,” a speech recognition of 6-10 corresponds to “GOOD,” a speech recognition of 11-15 corresponds to “FAIR,” and a speech recognition of 16-20 corresponds to “POOR”). In other examples, the speech recognition suitability can be mapped non-linearly to the different descriptions.

[0051] In other examples, the visual representation can include a numerical value representing the value of the speech recognition suitability. For example, FIG. 6 illustrates interface 600 that can be displayed at block 306 of process 300. As shown, interface 600 includes visual representation 602 in the form of the numerical value “96%.” In this example, visual representation 602 can include the numerical value of the speech recognition suitability or can include a normalized value of the speech recognition suitability (e.g., by normalizing the value to a percentage based on the speech recognition suitability value and the maximum possible speech recognition suitability value). For instance, if the speech recognition suitability values range from 0-25, a speech recognition suitability value of 1 can be linearly normalized to 96%. In other examples, the speech recognition suitability can be non-linearly normalized to different values.

[0052] In yet other examples, the visual representation can include an icon, and the value of the speech recognition suitability can be represented by a color of the icon. For example, FIG. 7 illustrates interface 700 that can be displayed at block 306 of process 300. As shown, interface 700 includes visual representation 702 in the form of an icon having a

microphone, where the value of the speech recognition suitability is represented by a color of the icon. In this example, the speech recognition suitability can be mapped to one of four (or any other desired number) different colors (e.g., grey, red, yellow, green) in which visual representation 702 can be displayed, with green representing the most suitable acoustic environment for performing speech recognition, yellow representing the second most suitable acoustic environment for performing speech recognition, grey representing the least suitable acoustic environment for performing speech recognition, and red representing the second least suitable acoustic environment for performing speech recognition. For example, if the speech recognition suitability values range from 1-20, the speech recognition suitability can be mapped linearly to the four discrete colors (e.g., a speech recognition of 1-5 corresponds to green, a speech recognition of 6-10 corresponds to yellow, a speech recognition of 11-15 corresponds to red, and a speech recognition of 16-20 corresponds to grey). In other examples, the speech recognition suitability can be mapped non-linearly to the different colors.

[0053] In some examples, the visual representation of the speech recognition suitability displayed at block 306 can be adjusted based on determined characteristics of the user. For example, the threshold values for determining the number of bars to display in interface 400, determining the text description to display in interface 500, determining the numerical value to display in interface 600, or determining the color to display in interface 700 can be adjusted based on the expected difficulty of performing speech recognition on speech produced by a user having certain characteristics. For example, user's that are female, that have an accent, and that are non-native speakers, can typically be more difficult for speech recognition programs to interpret. Thus, if the user has any of these characteristics, the user device can adjust the threshold values for determining the appropriate visual representation to require a better speech recognition suitability (indicating a more suitable acoustic environment) to display a particular visual representation. For example, a speech recognition suitability of 9 can result in four bars being filled-in in visual representation 402 for a male speaker, while a speech recognition suitability of 9 can result in only three bars being filled-in in visual representation 402 for a female speaker.

[0054] Process 300 can be repeated any number of times at any interval of time to provide update to date information regarding the suitability of an acoustic environment for performing speech recognition. For example, block 302 can be repeatedly performed to provide the user device with current audio input. Block 304 can then be performed to update

the speech recognition suitability and block 306 can be performed to update the visual representation of the determined speech recognition suitability values.

[0055] It should be appreciated that process 300 can be performed in response to user input, such as a user selecting visual representation 702, the user uttering a trigger phrase, or any other desired input, or can be performed during any other duration of time. For example, process 300 can be performed while user device 300 is turned on, while user device 300 is plugged in to a power source, while an option for determining suitability is selected, or the like. In these examples, process 300 can be repeatedly performed to display a visual representation of speech recognition suitability before, during, or after the user interacts with the user device in natural language.

[0056] It should further be appreciated that the blocks of process 300 can be performed on user device 102, server system 110, or a combination of user device 102 and server system 110. For instance, in some examples, all blocks of process 300 can be performed on user device 102. In other examples, some blocks of process 300 can be performed at user device 102, while other blocks of process 300 can be performed at server system 110. For example, blocks 302 can be performed at user device 102. User device 102 can then transmit the audio input to server system 110, which can then determine the speech recognition suitability based on the audio input at block 304. Server system 110 can transmit the determined speech recognition suitability to user device 110, which can then display the visual representation of the speech recognition suitability at block 306.

[0057] Using process 300, a user device can advantageously provide an indication of the suitability of an acoustic environment for performing speech recognition. This allows a user to determine whether to proceed with the performance of a speech recognition process, or to move to a different location having a better acoustic environment before performing the speech recognition process. Additionally, this can provide a user with useful information about an acoustic environment when performing speech recognition that would otherwise be unavailable. For example, if executed while performing speech recognition, process 300 can provide a user with an explanation for why the speech recognition process is producing erroneous results.

[0058] In some examples, visual representation 702 shown in FIG. 7 can be selectable by a user for the purpose of triggering a virtual assistant executed by the user device. For example, a user selection of visual representation 702 can be interpreted by the user device as

a start-point of user speech, causing the user device to monitor subsequently received audio input for user speech and performing speech recognition on any such identified speech. In some of these examples, the user device can disable the functionality of visual representation 702 in response to the speech recognition suitability determined at block 304 being below (alternatively, above, depending on the scoring convention used for the speech recognition suitability) a threshold value, indicating a poor acoustic environment for performing speech recognition. FIG. 8 illustrates an exemplary process 800 for operating a virtual assistant according to various examples. In some examples, process 800 can be performed after the visual representation of the speech recognition suitability has been displayed at block 306 of process 300. At block 802, it can be determined whether a selection of the visual representation (e.g., the icon of visual representation 702) has been received. In some examples, this can include determining whether a tap on a touch sensitive display of the user device has been received at a location corresponding to the visual representation. In other examples, other forms of user input can be monitored to determine if a selection of the visual representation has been received. If it is determined that no selection of the visual representation has been received, process 800 can repeat block 802 until a selection is received. Alternatively, if it is instead determined at block 802 that a selection of the visual representation has been received, process 800 can proceed to block 804.

[0059] At block 804, it can be determined whether the visual representation has been disabled due to a poor speech recognition suitability being determined for the acoustic environment at block 304 of process 300. In some examples, this can include determining whether the icon of visual representation 702 is displayed in a grey color or greyed out state. In other examples, the operation of the virtual assistant can be disabled in response to the speech recognition suitability being below a value that is different than the value used to delineate between display colors of the visual representation 702. In these examples, block 804 can include determining whether another condition is satisfied, such as determining whether the speech recognition suitability is less than a threshold value. If it is determined at block 804 that the visual representation is not disabled, process 800 can proceed to block 806 where a speech recognition process can be performed on audio input that is subsequently received by the user device. Alternatively, if it is instead determined at block 804 that the visual representation is disabled, process 800 can proceed to block 808.

[0060] At block 808, rather than perform speech recognition on audio input subsequently received by the user device, the user device can output a notification to the user

that indicates the detection of an acoustic environment that is not suitable for performing speech recognition. This notification can include any desired audio, visual, or haptic output. For example, the notification can include a text message displayed on a display of the user device instructing the user to move to a new location to perform the speech recognition process.

[0061] Using process 800, a user device can advantageously disable functionality of a speech recognition process in response to determining that the speech recognition suitability of the present acoustic environment is below a threshold suitability. This can prevent frustration by a user attempting to use the speech recognition process in environments where it is highly unlikely for the speech recognition to function properly.

ELECTRONIC DEVICE

[0062] In accordance with some examples, FIG. 9 shows a functional block diagram of an electronic device 900 configured in accordance with the principles of the various described examples. The functional blocks of the device can be 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. 9 can be combined or separated into sub-blocks to implement the principles of the various described examples. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0063] As shown in FIG. 9, electronic device 900 can include a touch screen display unit 902 configured to display a user interface and to receive touch input, and a sound receiving unit 904 configured to receive sound input. In some examples, electronic device 900 can include a speaker unit 906 configured to generate sound. Electronic device 900 can further include a processing unit 908 coupled to touch screen display unit 902 and sound receiving unit 904 (and, optionally, coupled to speaker unit 906). In some examples, processing unit 908 can include determining unit 910, displaying unit 912, selection determining unit 914, speech recognition unit 916, and message outputting unit 918.

[0064] Processing unit 908 can be configured to receive an audio input (e.g., from audio receiving unit 904). Determining unit 910 can be configured to determine a speech recognition suitability based on the audio input, wherein the speech recognition suitability represents a suitability of an acoustic environment of the electronic device for speech

recognition. Displaying unit 912 can be configured to display, in accordance with a determination of the speech recognition suitability, a visual representation of the speech recognition suitability.

[0065] In some examples, determining the speech recognition suitability based on the audio input includes: determining one or more characteristics of the acoustic environment based on the audio input; and determining the speech recognition suitability based on the one or more characteristics of the acoustic environment.

[0066] In some examples, the one or more characteristics of the acoustic environment includes a signal to noise ratio for a first frequency band of the acoustic environment. In some examples, the one or more characteristics of the acoustic environment includes a type of noise detected in the first frequency band.

[0067] In some examples, the one or more characteristics of the acoustic environment includes a signal to noise ratio for a second frequency band of the acoustic environment. In some examples, the one or more characteristics of the acoustic environment includes a type of noise detected in the second frequency band.

[0068] In some examples, the one or more characteristics of the acoustic environment includes a number of transient noises detected in a buffer comprising previously recorded audio of the acoustic environment.

[0069] In some examples, determining the speech recognition suitability based on the audio input includes: determining a speech recognition suitability vector based on the audio input, wherein the speech recognition suitability vector comprises one or more elements that represent the one or more characteristics of the acoustic environment; and using a neural network to determine the speech recognition suitability based on the speech recognition suitability vector.

[0070] In some examples, the visual representation includes one or more bars, and wherein a value of the speech recognition suitability is represented by a number of the one or more bars.

[0071] In some examples, the visual representation includes an icon, and wherein a value of the speech recognition suitability is represented by a color of the icon. In some examples, the icon includes an image of a microphone. In some examples, displaying the visual representation of the speech recognition suitability includes: determining whether a value of the speech recognition suitability is less than a threshold value; in accordance with a

determination that the value of the speech recognition suitability is less than the threshold value, displaying the icon in a grayed out state; and in accordance with a determination that the value of the speech recognition suitability is not less than the threshold value, displaying the icon in a non-grayed out state.

[0072] In some examples, selection determining unit 914 can be configured to determine whether a user selection of the icon is received. Speech recognition unit 916 can be configured to perform, in accordance with a determination that the user selection of the icon is received while the icon is displayed in the non-grayed out state, speech recognition on an audio input received subsequent to receiving the user selection of the icon. Speech recognition unit 916 can be further configured to forgo, in accordance with a determination that the user selection of the icon is received while the icon is displayed in the grayed out state, the performance of speech recognition on the audio input received subsequent to receiving the user selection of the icon.

[0073] In some examples, message outputting unit 918 can be configured to output, in accordance with a determination that the value of the speech recognition suitability is less than the threshold value, a message indicating a low suitability of the acoustic environment of the electronic device for speech recognition.

[0074] In some examples, the visual representation includes a textual representation of the speech recognition suitability.

[0075] In some examples, determining the speech recognition suitability based on the audio input includes periodically determining the speech recognition suitability based on the audio input, and displaying the visual representation of the speech recognition suitability includes updating the display of the visual representation of the speech recognition suitability in accordance with the periodically determined speech recognition suitability.

[0076] In some examples, the speech recognition suitability comprises a numerical value.

[0077] As described above, one aspect of the present technology is the gathering and use of data available from various sources to improve the delivery to users of invitational content or any other content that may be of interest to them. The present disclosure contemplates that in some instances, this gathered data can include personal information data that uniquely identifies or can be used to contact or locate a specific person. Such personal information

data can include demographic data, location-based data, telephone numbers, email addresses, home addresses, or any other identifying information.

[0078] The present disclosure recognizes that the use of such personal information data, in the present technology, can be used to the benefit of users. For example, the personal information data can be used to deliver targeted content that is of greater interest to the user. Accordingly, use of such personal information data enables calculated control of the delivered content. Further, other uses for personal information data that benefit the user are also contemplated by the present disclosure.

[0079] The present disclosure further contemplates that the entities responsible for the collection, analysis, disclosure, transfer, storage, or other use of such personal information data will comply with well-established privacy policies and/or privacy practices. In particular, such entities should implement and consistently use privacy policies and practices that are generally recognized as meeting or exceeding industry or governmental requirements for maintaining personal information data private and secure. For example, personal information from users should be collected for legitimate and reasonable uses of the entity and not shared or sold outside of those legitimate uses. Further, such collection should occur only after receiving the informed consent of the users. Additionally, such entities would take any needed steps for safeguarding and securing access to such personal information data and ensuring that others with access to the personal information data adhere to their privacy policies and procedures. Further, such entities can subject themselves to evaluation by third parties to certify their adherence to widely accepted privacy policies and practices.

[0080] Despite the foregoing, the present disclosure also contemplates examples in which users selectively block the use of, or access to, personal information data. That is, the present disclosure contemplates that hardware and/or software elements can be provided to prevent or block access to such personal information data. For example, in the case of advertisement delivery services, the present technology can be configured to allow users to select to “opt in” or “opt out” of participation in the collection of personal information data during registration for services. In another example, users can select not to provide location information for targeted content delivery services. In yet another example, users can select to not provide precise location information, but permit the transfer of location zone information.

[0081] Therefore, although the present disclosure broadly covers use of personal information data to implement one or more various disclosed examples, the present disclosure also contemplates that the various examples can also be implemented without the need for accessing such personal information data. That is, the various examples of the present technology are not rendered inoperable due to the lack of all or a portion of such personal information data. For example, content can be selected and delivered to users by inferring preferences based on non-personal information data or a bare minimum amount of personal information, such as the content being requested by the device associated with a user, other non-personal information available to the content delivery services, or publicly available information.

[0082] Although 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 various examples as defined by the appended claims.

WHAT IS CLAIMED IS:

1. A method for operating a virtual assistant, the method comprising:
at an electronic device:
receiving an audio input;
determining a speech recognition suitability based on the audio input, wherein the speech recognition suitability represents a suitability of an acoustic environment of the electronic device for speech recognition; and
in accordance with a determination of the speech recognition suitability, displaying a visual representation of the speech recognition suitability.
2. The method of claim 1, wherein determining the speech recognition suitability based on the audio input comprises:
determining one or more characteristics of the acoustic environment based on the audio input; and
determining the speech recognition suitability based on the one or more characteristics of the acoustic environment.
3. The method of claim 2, wherein the one or more characteristics of the acoustic environment comprises a signal to noise ratio for a first frequency band of the acoustic environment.
4. The method of claim 3, wherein the one or more characteristics of the acoustic environment comprises a type of noise detected in the first frequency band.
5. The method of claim 2, wherein the one or more characteristics of the acoustic environment comprises a signal to noise ratio for a second frequency band of the acoustic environment.
6. The method of claim 5, wherein the one or more characteristics of the acoustic environment comprises a type of noise detected in the second frequency band.

7. The method of claim 2, wherein the one or more characteristics of the acoustic environment comprises a number of transient noises detected in a buffer comprising previously recorded audio of the acoustic environment.
8. The method of claim 2, wherein determining the speech recognition suitability based on the audio input comprises:
 - determining a speech recognition suitability vector based on the audio input, wherein the speech recognition suitability vector comprises one or more elements that represent the one or more characteristics of the acoustic environment; and
 - using a neural network to determine the speech recognition suitability based on the speech recognition suitability vector.
9. The method of claim 1, wherein the visual representation comprises one or more bars, and wherein a value of the speech recognition suitability is represented by a number of the one or more bars.
10. The method of claim 1, wherein the visual representation comprises an icon, and wherein a value of the speech recognition suitability is represented by a color of the icon.
11. The method of claim 10, wherein the icon comprises an image of a microphone.
12. The method of claim 10, wherein displaying the visual representation of the speech recognition suitability comprises:
 - determining whether a value of the speech recognition suitability is less than a threshold value;
 - in accordance with a determination that the value of the speech recognition suitability is less than the threshold value, displaying the icon in a grayed out state; and
 - in accordance with a determination that the value of the speech recognition suitability is not less than the threshold value, displaying the icon in a non-grayed out state.
13. The method of claim 12, wherein the method further comprises:
 - determining whether a user selection of the icon is received;

in accordance with a determination that the user selection of the icon is received while the icon is displayed in the non-grayed out state, performing speech recognition on an audio input received subsequent to receiving the user selection of the icon; and

in accordance with a determination that the user selection of the icon is received while the icon is displayed in the grayed out state, forgoing the performance of speech recognition on the audio input received subsequent to receiving the user selection of the icon.

14. The method of claim 12, wherein the method further comprises:

in accordance with a determination that the value of the speech recognition suitability is less than the threshold value, outputting a message indicating a low suitability of the acoustic environment of the electronic device for speech recognition.

15. The method of claim 1, wherein the visual representation comprises a textual representation of the speech recognition suitability.

16. The method of claim 1, wherein determining the speech recognition suitability based on the audio input comprises periodically determining the speech recognition suitability based on the audio input, and wherein displaying the visual representation of the speech recognition suitability comprises updating the display of the visual representation of the speech recognition suitability in accordance with the periodically determined speech recognition suitability.

17. The method of claim 1, wherein the speech recognition suitability comprises a numerical value.

18. A computer-readable storage medium comprising instructions for performing the method of any of claims 1-17.

19. A system comprising:

the computer readable storage medium of claim 18; and

one or more processors capable of executing the instructions of the non-transitory computer readable storage medium.

20. A system comprising:
- means for receiving an audio input;
 - means for determining a speech recognition suitability based on the audio input,
- wherein the speech recognition suitability represents a suitability of an acoustic environment of the electronic device for speech recognition; and
- means for displaying, in accordance with a determination of the speech recognition suitability, a visual representation of the speech recognition suitability.

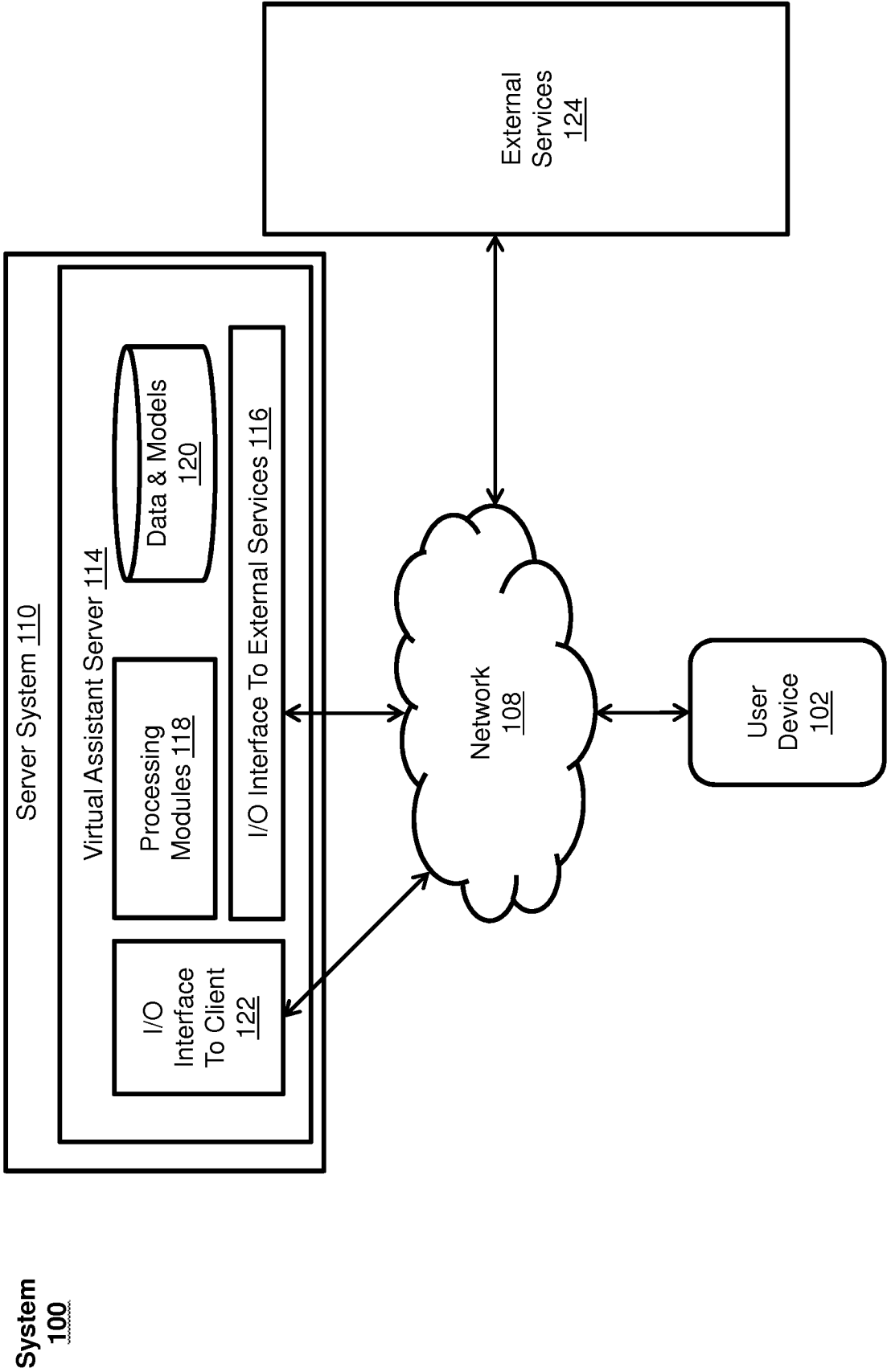


FIG. 1

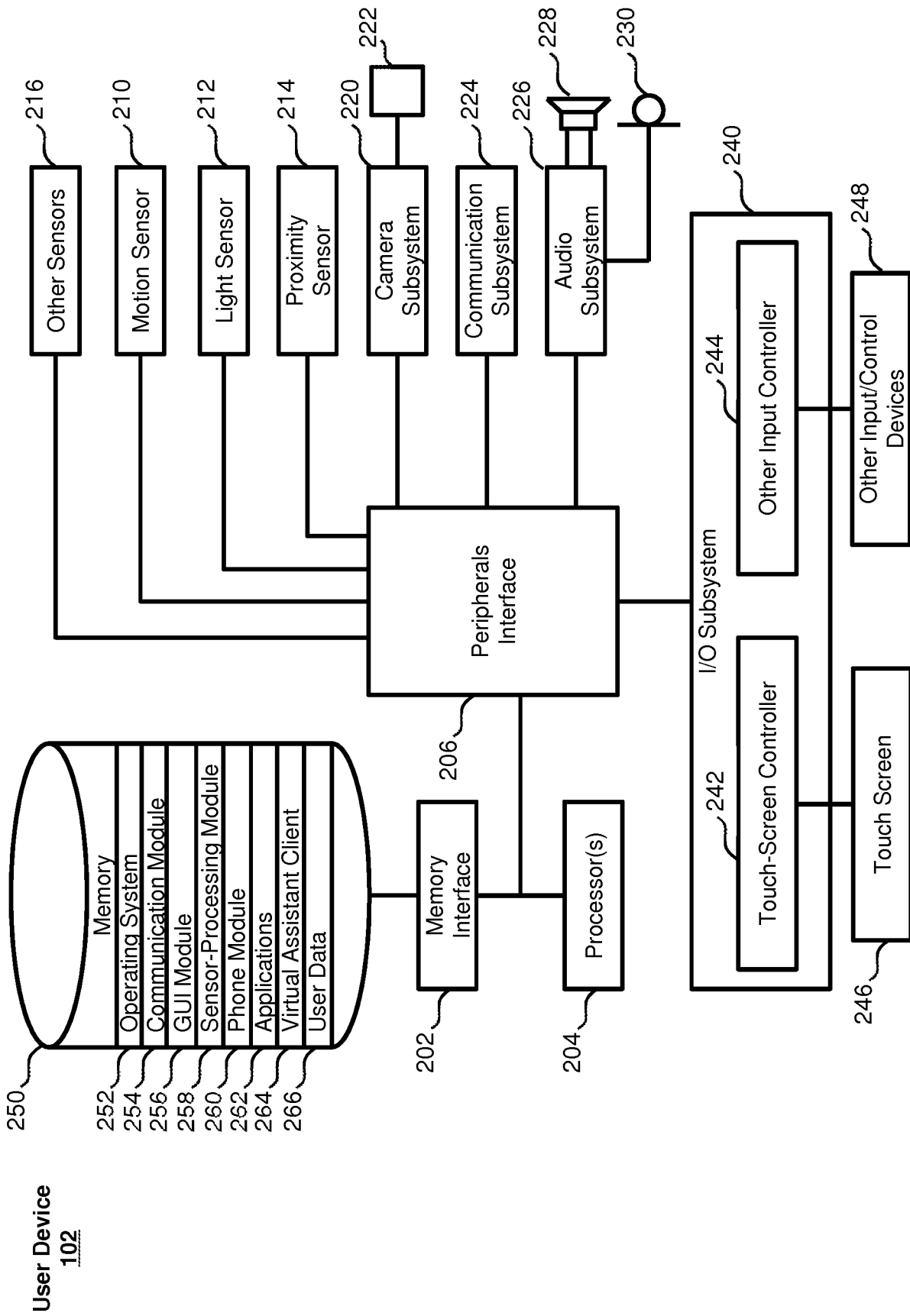


FIG. 2

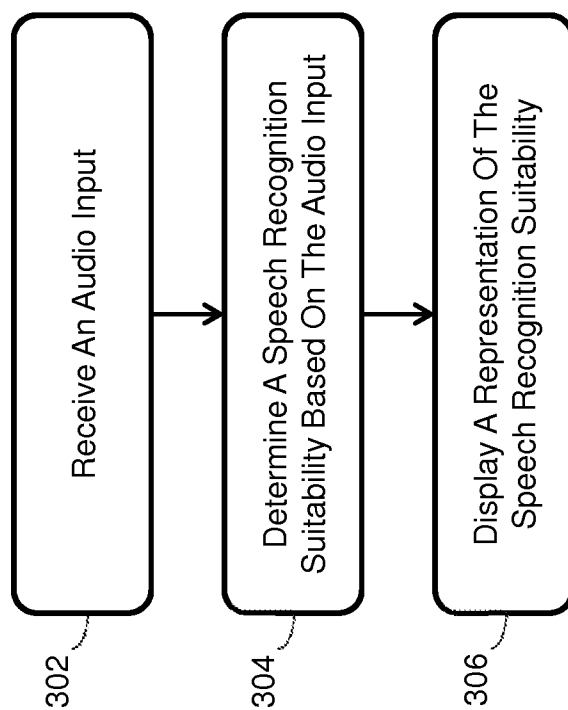


FIG. 3

Interface
400

Visual
Representation
402

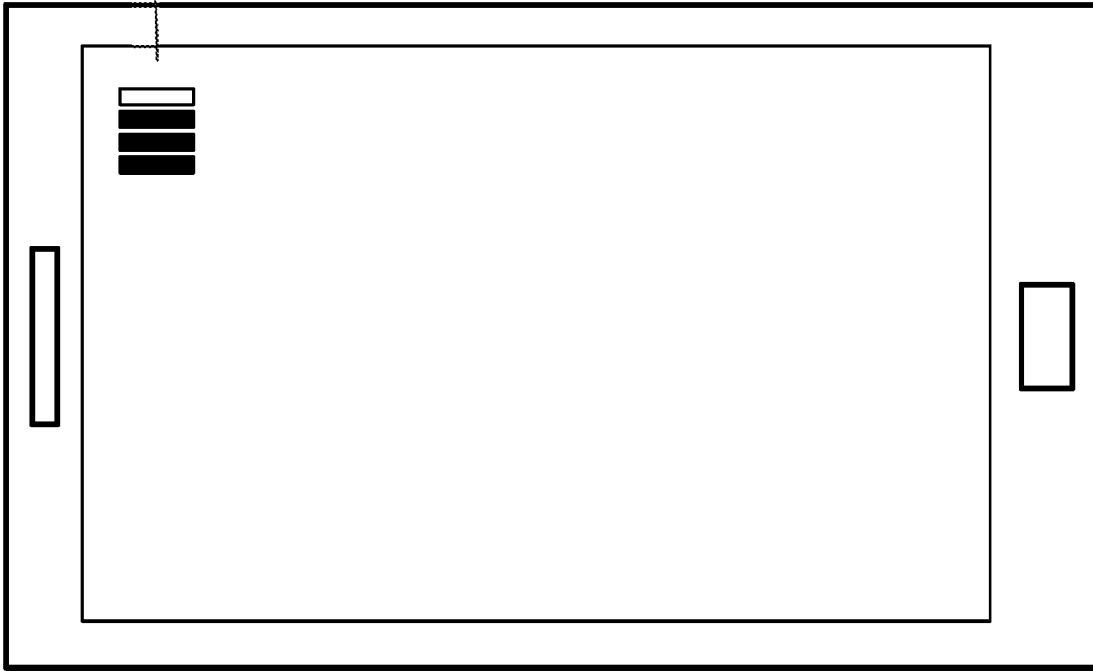


FIG. 4

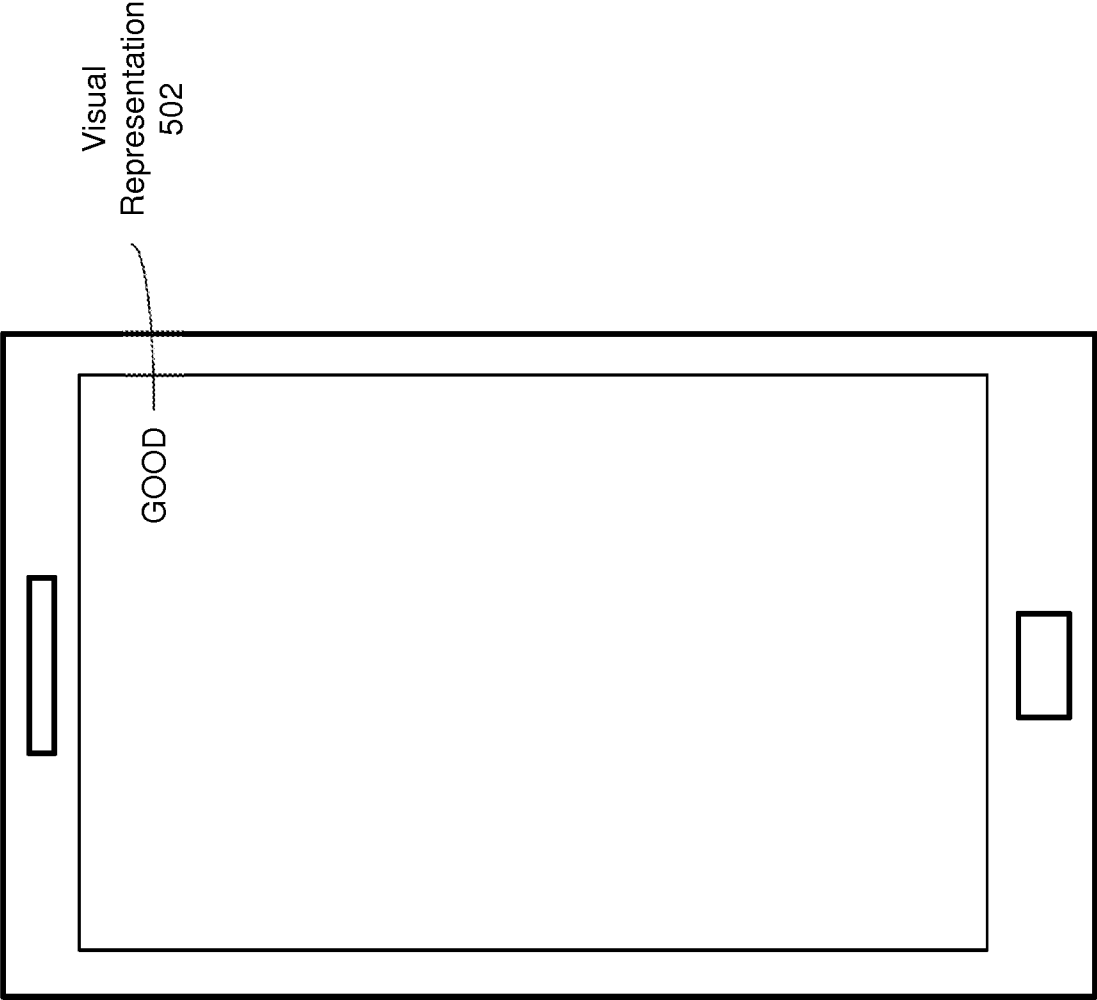


FIG. 5

Interface
500

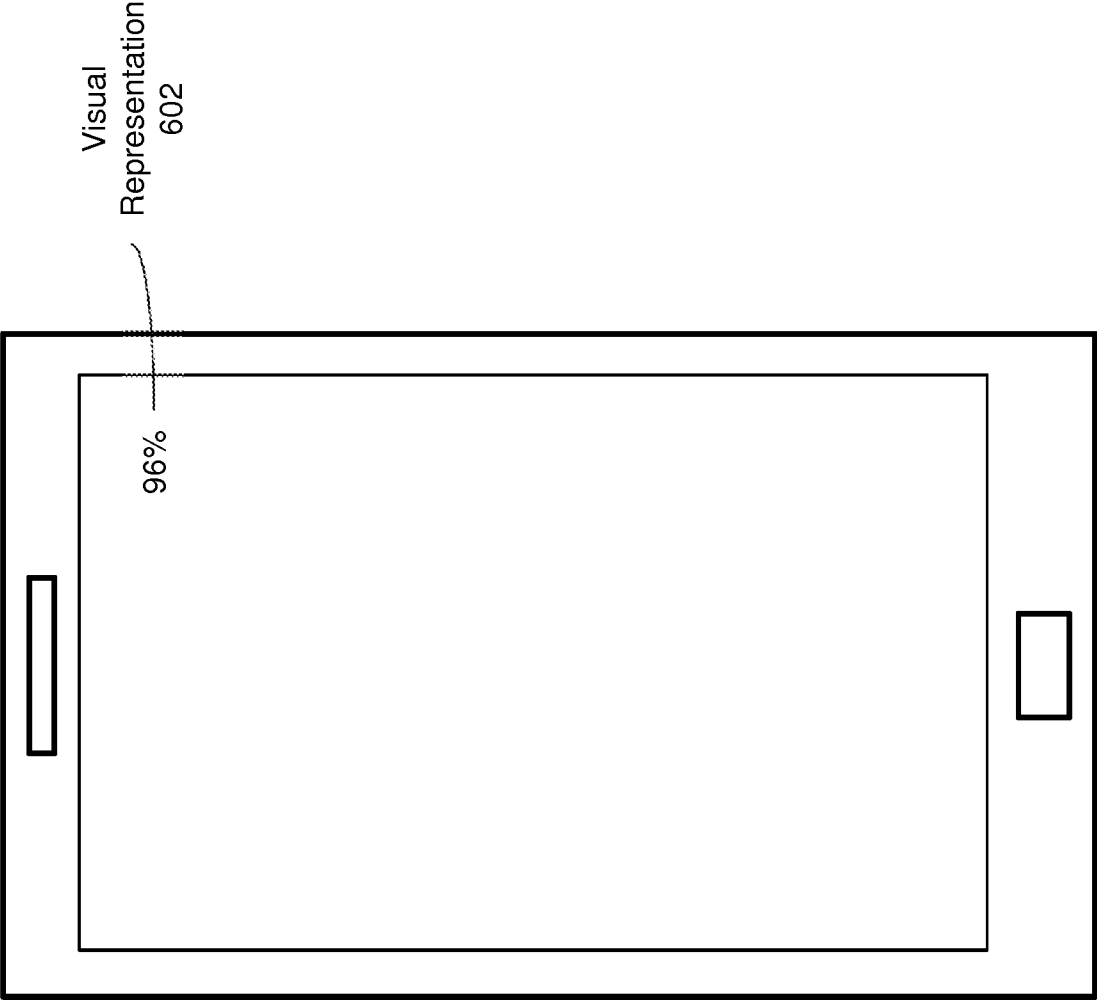


FIG. 6

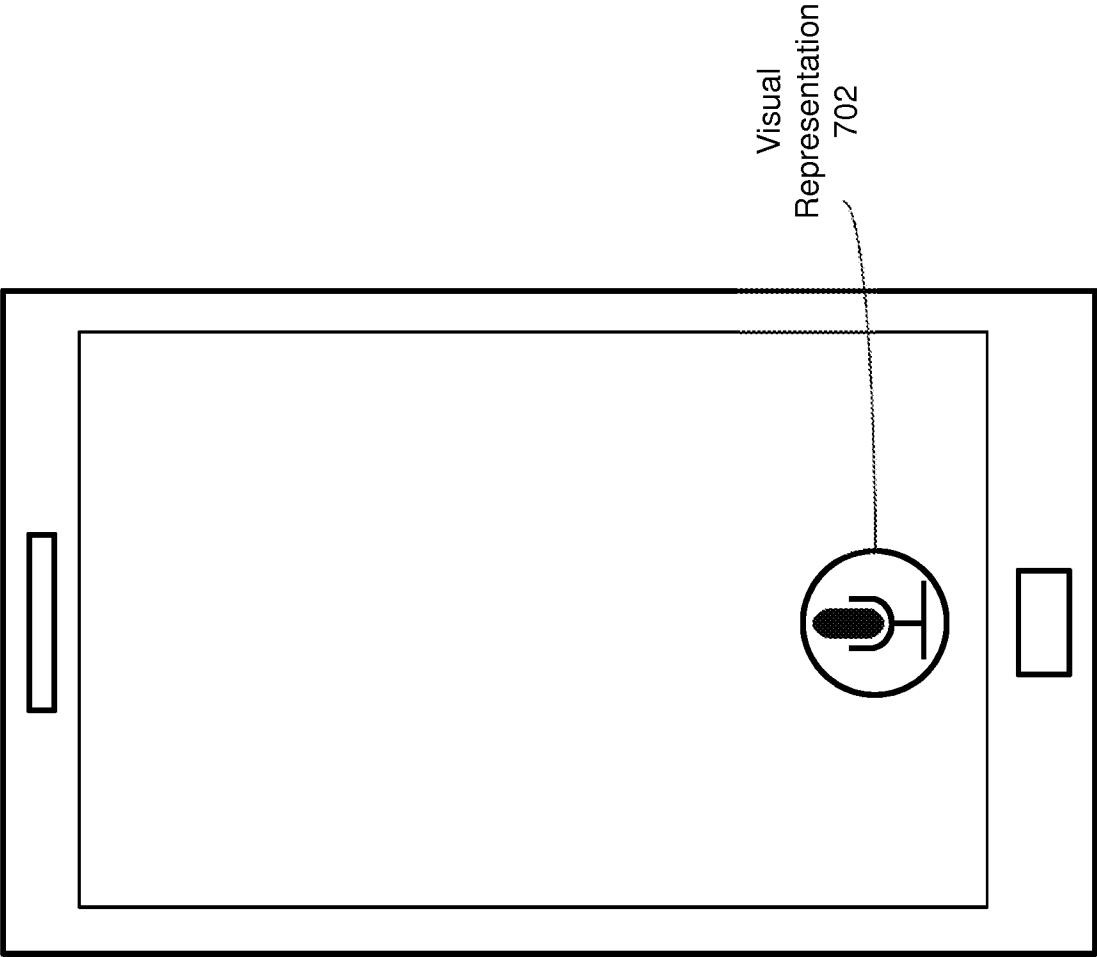


FIG. 7

Interface
700

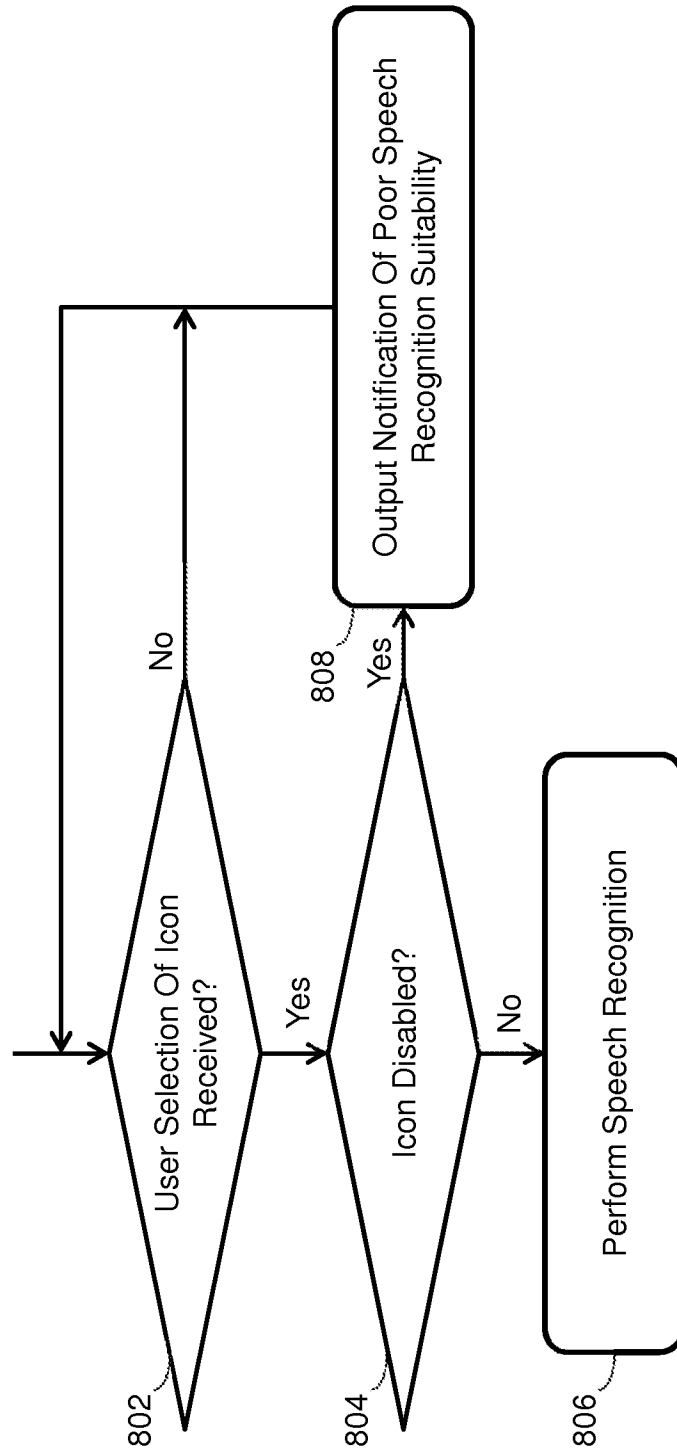


FIG. 8

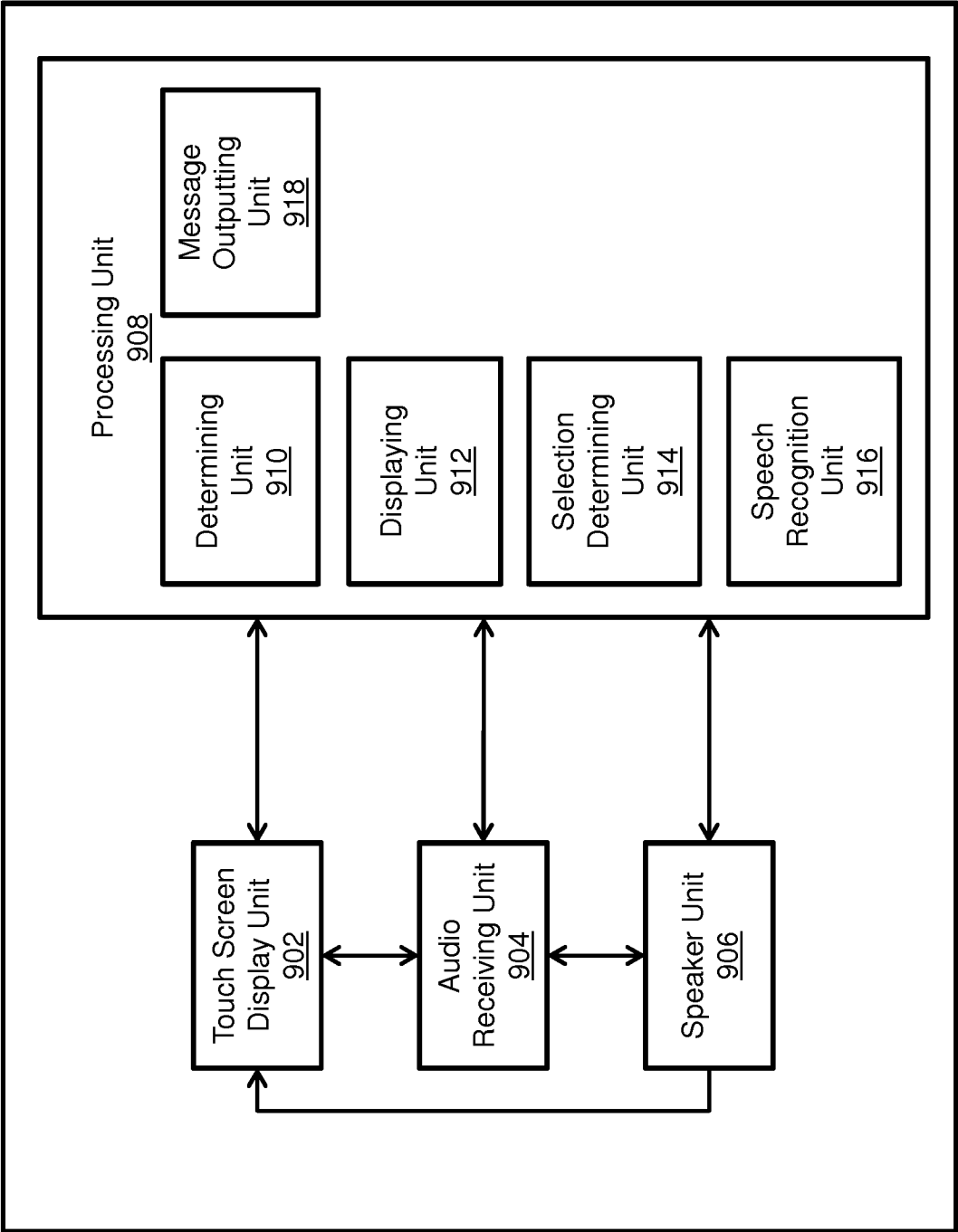


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/047553

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/22
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2008/162120 A1 (MACTAVISH THOMAS J [US] ET AL) 3 July 2008 (2008-07-03) paragraph [0011]; figure 4a -----	1-20
A	US 2013/158977 A1 (SENIOR ANDREW [US]) 20 June 2013 (2013-06-20) paragraph [0013] - paragraph [0030]; figures 1,2 -----	1-20
A	US 2002/196911 A1 (GAO YUQING [US] ET AL) 26 December 2002 (2002-12-26) paragraph [0057] - paragraph [0058]; claims 1, 5,6 -----	1-20
A	US 2013/041661 A1 (LEE WOO BEUM [US] ET AL) 14 February 2013 (2013-02-14) paragraph [0030] paragraph [0032] - paragraph [0033] paragraph [0035]; figure 7 -----	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

11 December 2015

Date of mailing of the international search report

05/01/2016

Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Perez Molina, E

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2015/047553

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 2008162120	A1	03-07-2008	NONE	
US 2013158977	A1	20-06-2013	NONE	
US 2002196911	A1	26-12-2002	NONE	
US 2013041661	A1	14-02-2013	NONE	