



(12) **United States Patent**  
**Mai**

(10) **Patent No.:** **US 10,002,259 B1**  
(45) **Date of Patent:** **Jun. 19, 2018**

(54) **INFORMATION SECURITY/PRIVACY IN AN ALWAYS LISTENING ASSISTANT DEVICE**

2014/0119531 A1 5/2014 Tuchman et al.  
2014/0372109 A1 12/2014 Iyer  
2015/0221307 A1\* 8/2015 Shah ..... G10L 15/28  
704/253

(71) Applicant: **Xiao Ming Mai**, East Brunswick, NJ (US)

2015/0279366 A1 10/2015 Krestnikov et al.  
2016/0077794 A1 3/2016 Kim et al.  
2016/0253710 A1 9/2016 Publicover et al.  
2017/0148444 A1 5/2017 Booklet et al.  
2017/0162205 A1 6/2017 Melvin et al.

(72) Inventor: **Xiao Ming Mai**, East Brunswick, NJ (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days. days.

#### FOREIGN PATENT DOCUMENTS

CN 102893327 A 1/2013  
CN 20141784875 A 7/2016  
WO 1996002911 A1 2/1996  
WO 2007121570 A1 11/2007  
WO 2015148693 A1 3/2014

\* cited by examiner

(21) Appl. No.: **15/812,679**

(22) Filed: **Nov. 14, 2017**

(51) **Int. Cl.**  
**G06F 21/70** (2013.01)  
**G06F 21/60** (2013.01)  
**H04W 12/02** (2009.01)

*Primary Examiner* — Izunna Okeke  
(74) *Attorney, Agent, or Firm* — Weitzman Law Offices, LLC

(52) **U.S. Cl.**  
CPC ..... **G06F 21/604** (2013.01); **H04W 12/02** (2013.01)

#### (57) **ABSTRACT**

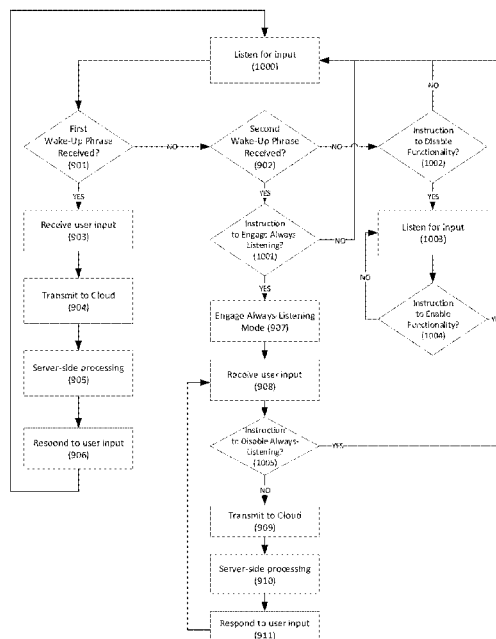
An always-listening-capable computing device includes a receiver for input from a user, a module for communication with a remote server, and a gate-keeping module that, when enabled, prevents the communication module from transmitting data external to the device. After determining that user input includes a first wake up phrase, the device processor automatically transmits a representation of user input subsequent to the phrase, activates an always-receiving mode to transmit a stream of user input captured subsequent to the phrase, deactivates the always-receiving mode to prevent transmission of user input received subsequent to the phrase, unless also preceded by a second wake up phrase, or enables the gate-keeping module to prevent transmission of data external to the device.

(58) **Field of Classification Search**  
CPC ..... G06F 1/3206; G06F 21/71; G06F 9/4806  
See application file for complete search history.

(56) **References Cited**  
**U.S. PATENT DOCUMENTS**

8,194,986 B2 6/2012 Conwell  
8,374,879 B2 2/2013 Falcon et al.  
8,655,307 B1 2/2014 Walker et al.  
8,990,235 B2 3/2015 King et al.  
9,338,493 B2 5/2016 Van Os et al.  
2010/0332236 A1 12/2010 Tan  
2011/0060588 A1 3/2011 Weinberg

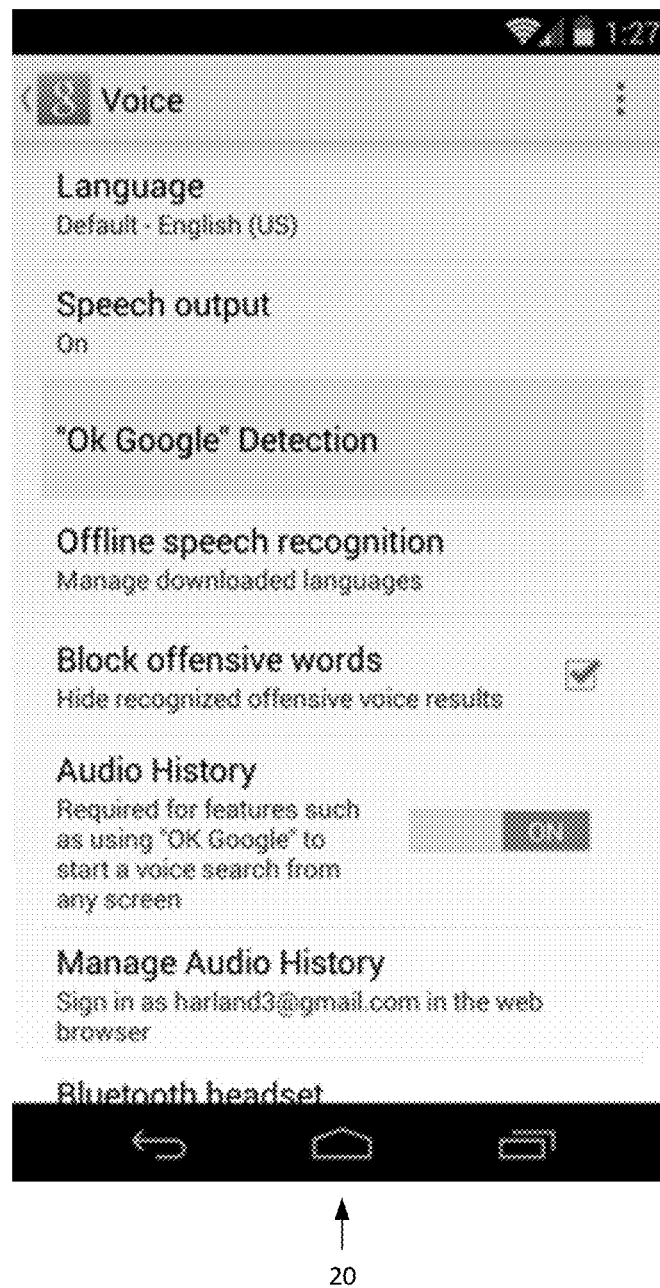
**17 Claims, 10 Drawing Sheets**





(PRIOR ART)

**Fig. 1**



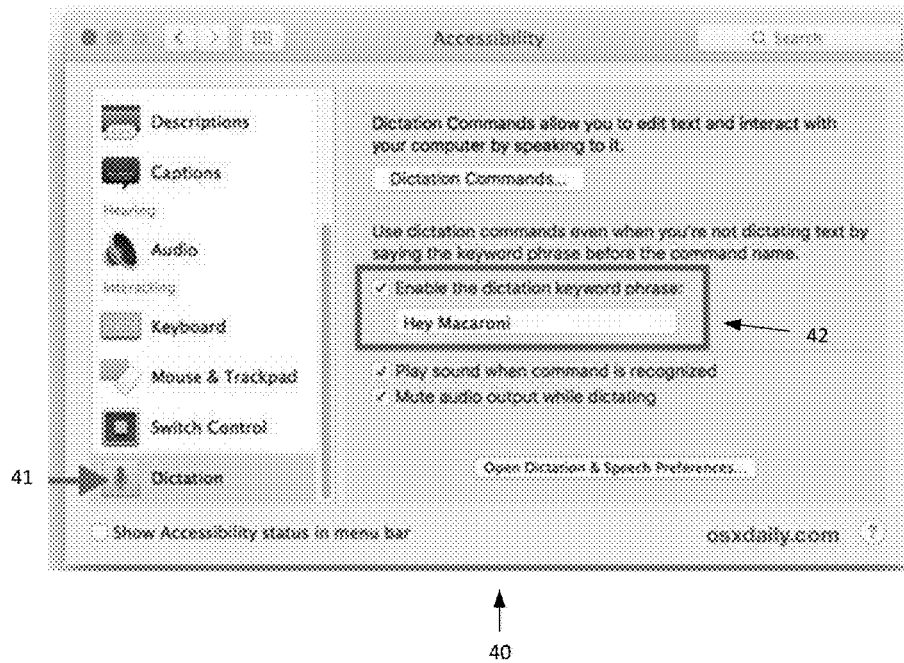
(PRIOR ART)

**Fig. 2**



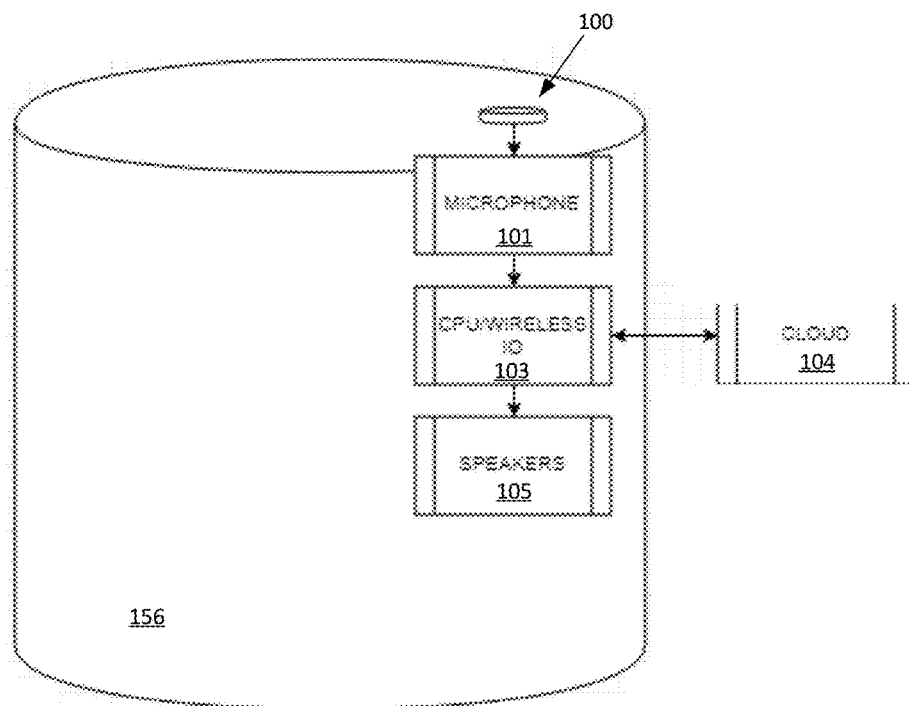
(PRIOR ART)

**Fig. 3**



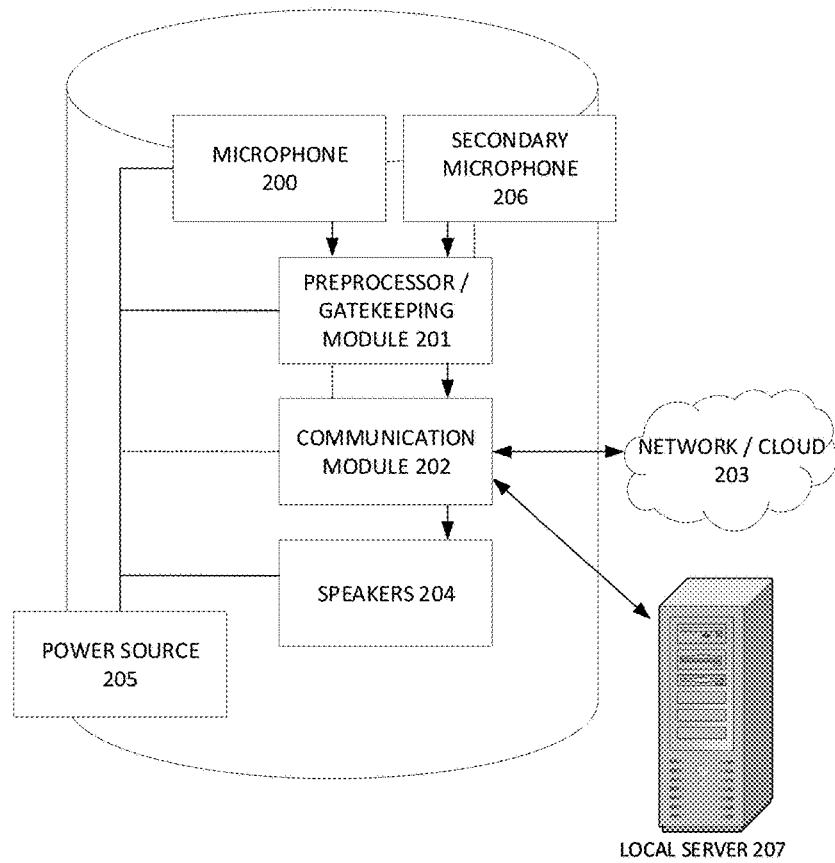
(PRIOR ART)

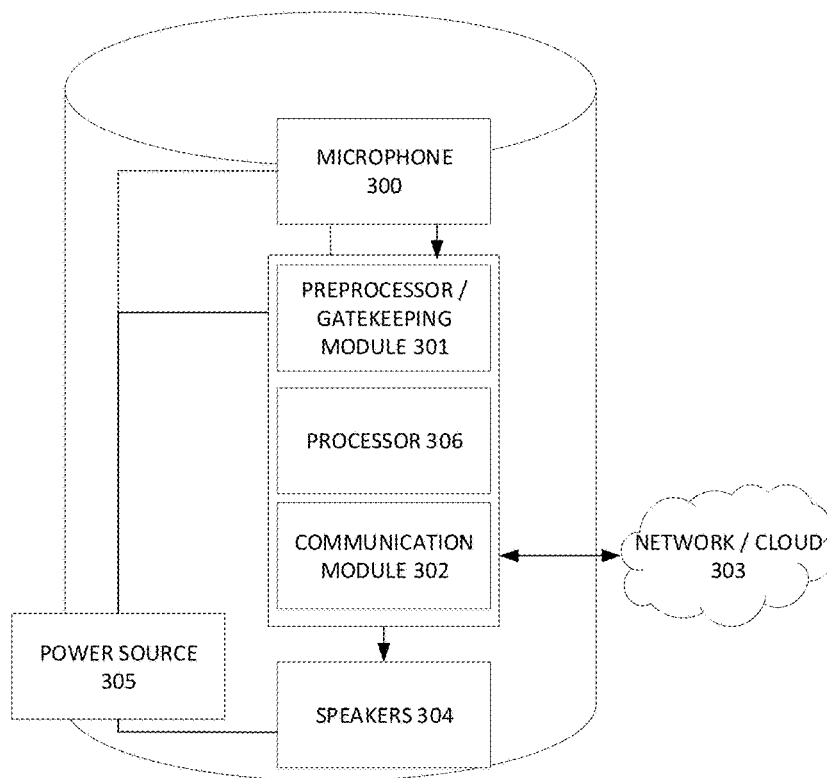
**Fig. 4**



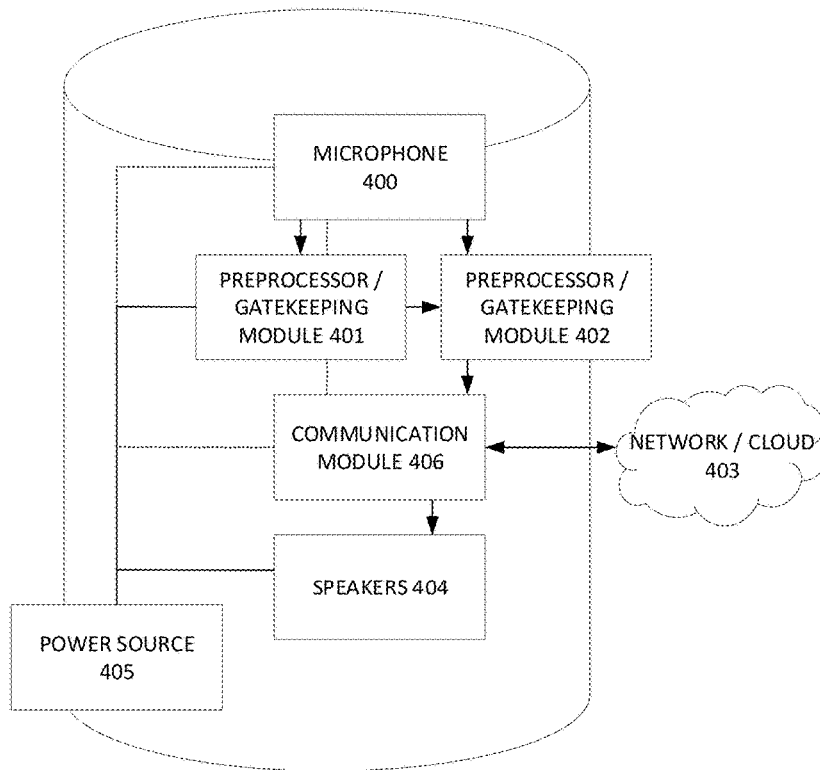
(PRIOR ART)

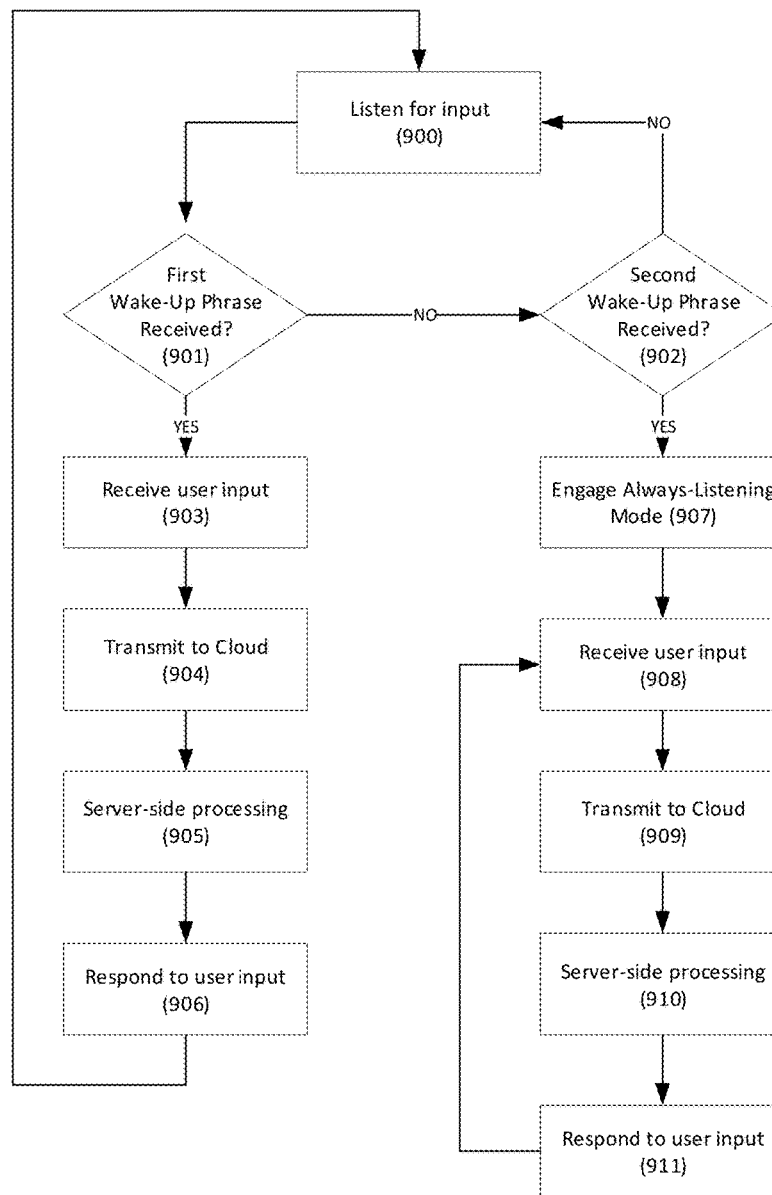
**Fig. 5**

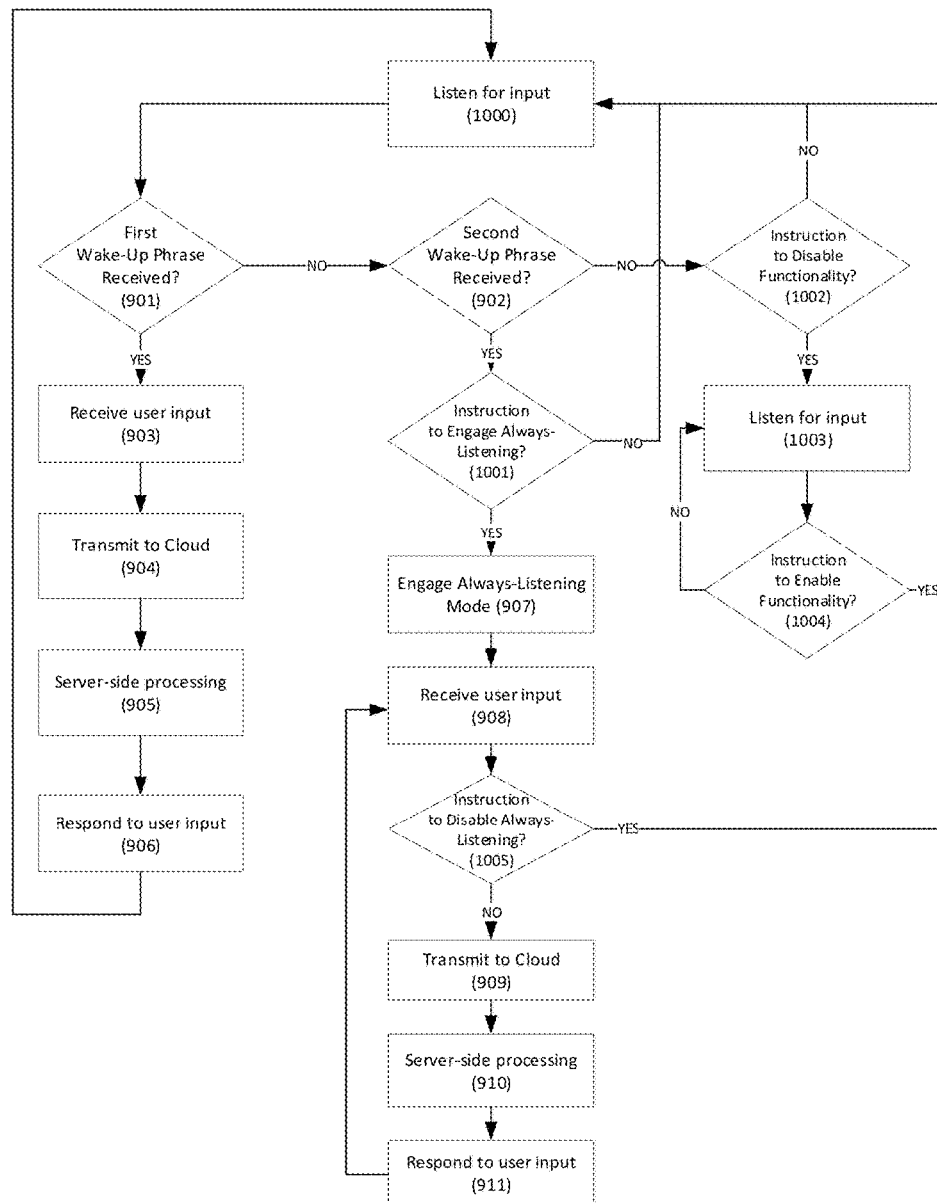
**Fig. 6**

**Fig. 7**



**Fig. 8**

**Fig. 9**

**Fig. 10**

1

## INFORMATION SECURITY/PRIVACY IN AN ALWAYS LISTENING ASSISTANT DEVICE

### FIELD OF INVENTION

This application relates to methods and systems for security and privacy of audible speech data, and performing an assistive response to the audible speech data via an “always-listening” medium.

### BACKGROUND

Many devices or software solutions are currently marketed to consumers as “always listening,” including Amazon Echo® (a.k.a. Alexa™), Google Home™, Apple HomePod™ and Siri® on Apple® devices. However, these devices are configured to always listen for their respective “wake up” words, upon receipt of which the devices will only continue to record additional speech for a period of time (e.g., a single complete command, string or query proceeding their wake up words, ending with a pause). These wake up words let the devices know when a command or query is directed to these devices such that the immediate proceeding speech data is captured and processed; and no other speech data is intended to be recorded or sent to the cloud in any way, shape or form, in order to safeguard the security and protect privacy of audible data spoken by users.

The current “always listening” devices would more appropriately be termed “always listening for a wake up word,” and are, in fact, inactive and dormant in response to most speech, rather than “always” analyzing and determining a response to received speech and commands.

Further, these devices only respond by complying with a single command or responding to a single query; they are unable to comply or intuitively follow a series of commands or queries due to the limitations of the current systems and methods of “listening.”

Current market-available solutions for protecting data use a physical button on the surface of the device to provide users complete control over the microphone and whether these devices can listen at all. Requiring a manual button to ensure enabling and disabling the microphone is inconvenient when consumers expect voice-only control, and is not a satisfactory solution. Device manufacturers primarily offer it for the users’ peace of mind.

Some systems control “always listening” status by software means, which are always at risk of external digital intruders hacking in and stealing audible or speech data from the microphone.

Consumers are unlikely to trust that these devices and security protocols will completely protect their privacy, especially if there is no direct speech control provided over the “always listening” function.

The current methods of using “wake up” words and security software protocols are time and labor intensive to constantly develop, maintain, and improve—while still failing to achieve guaranteed security of muting the device when a user so desires.

Most current technological focus has been emphasized on improving the mechanics or the “how-to’s” of detecting audible speech, improving accuracy and offering “faster responses”—where the latter has been focused on pre-defining rules and programs to respond to queries or pre-analyzing past behaviors which is extremely limited given the vast possible variety of queries and commands or desired assistance that millions of users would have, each more than likely to be unique.

2

Device makers and other advertisers compete to balance delivering the most desired assistance at the most applicable timing, subject relevancy, based on user need and receptiveness, with being perceived by the user as non-intrusive and non-disruptive.

### SUMMARY OF THE INVENTION

Disclosed herein is an improved technical method, device, and system for users to have complete control over when their speech or actions (including, but not limited to, words, tones, command gestures, communications in sign language, facial expressions, patterns of movement, or any combinations thereof) are recorded by electronic devices, while minimizing or eliminating the need for security programming, organically creating a system of controlled access to electronic device and audible speech and/or video data, and offering both users and assistance delivery service providers a channel to receive, analyze, decipher, compute, access, or assess data to deliver real time relevant assistance to the user(s).

The present disclosure provides substantial improvements over current technologies and methods of securing received data as is further described herein towards providing users a means of using speech or gesture based actions to control an always listening mode in a listening medium or device (e.g., Amazon Echo®, mobile device, computing device, webcam, security camera system, etc.) without dependency on software to secure, therefore improving the acceptability to users to elect to allow the medium to always listen to conversations; feeding these directed or indirect speech, actions/gestures and/or commands/queries in near-real time to be deciphered, processed and provide undirected and unsolicited assistance or assistance without the need of a “wake up” word when full always-listening mode is active. Gesture(s) include (but are not limited to) custom gestures or language-based gestures such as, for example, American Sign Language.

In one aspect of this disclosure, an always-listening-capable computing device is disclosed herein. The device includes a first receiver of one or more receivers, configured to receive input from a user comprising an utterance or a gesture. The device also includes a communication module configured to transmit data to and receive data from a remote server and one or more processors. The device further includes a gate-keeping module that, when enabled, is configured to prevent the communication module from transmitting data external to the always-listening-capable computing device. Finally, the device includes non-transient memory storing instructions that, when executed by at least one processor, causes the at least one processor to determine that the input received by the first receiver includes a first wake up phrase. In response to determining that the input comprises the first wake up phrase, the at least one processor automatically performs at least one of: transmitting a representation of user input subsequent to the first wake up phrase; activating an always-receiving mode to transmit a stream of user input captured subsequent to the first wake up phrase; deactivating the always-receiving mode to prevent transmission of user input received subsequent to the first wake up phrase, unless also preceded by a second wake up phrase; or enabling the gate-keeping module to prevent all transmission of data external to the always-listening-capable computing device.

In another aspect of this disclosure, a computer-implemented method for a device to provide always listening assistive response includes receiving permission, via a

locally provided user input, to enter an always listening mode. User input comprising one or more directed utterances or gestures and directed to a recipient other than the device is received, via a receiver. A desire of the user for an action or for information is determined, based on user input, and at least one of the following are performed: an action that satisfies the user desire, or outputting, via a speaker or display, information that satisfies the user desire.

In another aspect of this disclosure, a method of providing information security in a data-recording system includes enabling, for a first interval of time, a wireless connection to receive data. The wireless connection is disabled, for a second interval of time shorter than the first interval of time, from receiving all incoming wireless data. The wireless connection is re-enabled to receive incoming wireless data.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts software controls in the prior art for the purpose of muting the function for “always listening” for the wake up words “Hey Siri” on an Apple® iPhone®;

FIG. 2 depicts software controls in the prior art for the purpose of muting the function for “always listening” for the wake up words “Ok Google” on an Android® OS device;

FIG. 3 depicts software controls in the prior art for the purpose of muting the microphone on a Windows® 10 OS;

FIG. 4 depicts third party software’s prior art functionality for setting a wake up word to command the software to begin dictation;

FIG. 5 depicts, in simplified form, the conventional relationship of components within a prior art personal digital assistant;

FIG. 6 depicts, in simplified form, an embodiment of an “always listening”-capable electronic device;

FIG. 7 depicts, in simplified form, an alternative embodiment of an “always listening”-capable electronic device;

FIG. 8 depicts, in simplified form, an alternative embodiment of an “always listening”-capable electronic device; and

FIG. 9 depicts a flow chart of an illustrative method of processing a second, distinct wake up word.

FIG. 10 depicts an alternative flowchart of program logic for additional features of an always-listening device.

### DETAILED DESCRIPTION

An improved method, apparatus and system for protecting audible data are disclosed herein, including a microphone or camera (herein referenced as “receiver”) incapable of receiving electrical signals (remaining capable of transmitting electrical signal and data), and may be one or more independent processor(s) incapable of receiving external electrical signal (while capable of receiving data) or gate keeping module embedded on a bi-directional processor where the gate keeping module resides where the gate keeping module is incapable of receiving external electrical signals (remaining capable of transmitting electrical signal and receiving data from a receiver), where the gate keeping module or independent processor (incapable of receiving external electrical signals) shall be act as a localized pre-processor to preprocess a user’s voice command prior to enable or disable truly “always listening” of the electronic device. Localized, unhackable control to turn on and off true always-listening via voice command are thus provided. This preprocessor or preprocessing (or gate keeping) function may be capable of directly relaying audible data to a wireless communication module, or to the primary processor to perform other processing functions such as end-to-end

encryption before relaying the then encrypted audible data to a wireless communication module; which then may connect to a remote computer (e.g., cloud servers or computing platforms). In essence, the preprocessor/gate keeping module acts as the unhackable “valve” of one-way communication that feeds received data from the receiver to the primary processor or wireless communications component.

The methods described herein utilize a localized preprocessor or preprocessing function on a processor that is incapable of receiving external electrical signals, isolating the “always listening” function and control to a local user, where the control of the gate keeping module is literally without outside data access and thus not susceptible to hacking.

FIGS. 1-4 depict various software controls in the prior art for controlling the listening and response of a digital personal assistant device. For example, interface 10 of FIG. 1 shows a toggle 11 for allowing the wake up words “Hey Siri” to trigger listening without a physical interaction with the device. Similarly, FIG. 2 displays an Android® interface 20 allowing the words “OK Google” to trigger listening. FIG. 3 displays an interface 30 in a Windows® operating system with a button 31 for disabling microphone input to the system. Finally, FIG. 4 displays an interface 40 in third party software “Dragon® Dictation” for controlling dictation generally, in tab 41 and setting a particular wake up phrase 42 to trigger recording and action.

FIG. 5 depicts, in simplified form, the conventional relationship of components within a prior art personal digital assistant. A user may press a button 100 to physically disable (mute) the microphone 101 from “always listening for a wake up” word. Under normal operating conditions (un-muted), microphone 101 passively listens for a wake up word(s) to detect, and once detected, will record a speech string, query or command immediately following the precursor wake up words to be sent to the processor and/or wireless I/O component 103. Once the string is received by the processor or wireless I/O component 103, the singular string is in most cases transmitted to a cloud server 104, or processed locally if the processor stores a set of preprogrammed executable instructions such as, for example, “turn off <name of> light.” The processor/always listening device then returns to passively listening for the precursor wake up word(s). If the string is transmitted rather than processed locally, the cloud server 104 processes the string, query or command, and returns a response to the first processor/through the Wireless I/O component 103. The response is then processed to actually comply with a command, or play an audible response through the speakers 105.

FIG. 6 depicts, in simplified form, an embodiment of a truly “always listening”-capable electronic device. The device could alternatively be described in other embodiments as “always receiving,” “always watching,” “always recording,” “always transmitting,” and/or any other method of sensing and recording input from sensors, whether those sensors are microphones, cameras, or other capturing devices.

Microphone (or other receiver) 200 may be configured to always be listening (or recording, or monitoring sensory feeds other than audio input) for potential instructions from a user, including, but not limited to, activating an always listening mode, deactivating an always listening mode, activating or deactivating features of an underlying digital personal assistant system, or providing digital personal assistant services such as responding to verbal queries.

5

Recorded audio or other data may be transmitted from receiver **200** to a preprocessor and/or a gate keeping module **201**.

In some embodiments, the device may comprise two or more distinct receivers. The first receiver **200** may be powered on and monitoring for user input so long as the device is powered, but may have no data leads to the device's communications module **202**, and thus may be prevented from communication with remote servers or other outside devices. Data output from the first receiver may be sent only to a processor with limited language processing ability and/or a gatekeeping module **201**. The second receiver **206** may normally be powered down or cut off from data transmission, with power being restored or data transmission functionality being restored only when permitted by the gatekeeping module **201**.

The gate keeping module **201** may control or restrict the means of inbound (including but not limited to authorized and unauthorized electrical signals) or outbound data (audio or speech data) at a critical path, component or function of the electronic device. Alternatively, the gate keeping module **201** may permit or deny power from power source **205** (which may be, for example, a battery or a connection to A/C power) to particular components of the device, including (but not limited to) a microphone, processor, and/or communications module (e.g., Wi-Fi, ZigBee, Bluetooth, Near Field Communication (NFC), cellular phone connection, etc.). In some embodiments, the various components are always directly connected to power source **205** (indicated by solid lines in the figure) while others may be either directly connected or connected through the gatekeeping module **201** (indicated by dashed lines in the figure).

The gate keeping module **201** may be implemented as an integrated circuit, a chip on a motherboard, or any other form of hardware solution. Alternatively, the gate keeping module **201** may take on the embodiment of security programming or protocols that are stored on non-transient memories, such as (but not limited to) EPROM (a memory that can only be erased locally and physically by shining a UV light onto the a designed/designated area on the memory chip and cannot be reprogrammed remotely). It should be understood that a person of ordinary skill in the art would be able to apply the gate keeping method as described herein in any number of alternative hardware and/or software embodiments.

The programming/protocols may have a function to limit inbound electrical signals received to a preset size or period of time or based on the most recent activity (e.g., sending a request to a server to load a webpage might permit an electronic device to receive data for a period time or until the webpage is loaded, but when there is no recent data request within a preset period of time, the gate keeping module within the EPROM may prevent additional data to be received by the electronic device through the wireless receiver) to prevent the electronic device from being hacked.

Alternatively, a gate keeping module may be set to "block (or limit) continuous audio transmissions on wireless transmitter," if a gate keeping module-processor does not have local input permissions from a user to continuously transmit.

The gate keeping module may be set to permit microbursts of inbound data during time intervals (e.g., anywhere from picoseconds, to seconds, to minutes) to enable an expected stream of data to be downloaded by the device, to allow checking for push notifications, and to allow reception of responses by a server configured to process user data, but make difficult or impossible a continuous connection needed to gain unauthorized access or "hack" a security

6

loophole, upload malicious data, or perform a denial of service attack. The length of the intervals, or the length of periods of disabled communication between the intervals, may be selected randomly or at fixed intervals (such as, for example, disabling communication for one millisecond every twenty milliseconds, or for one randomly chosen millisecond out of every second). The length of intervals may also be related as a proportion (such as, for example, inbound data being allowed for five times as long, twenty times as long, or any other multiplier of time compared to the interval of time that inbound data is disabled).

The gate keeping module may be designed to allow incoming data connections to the function itself only locally by physical, wired means or by more limited-range wireless means such as (but not limited to) Bluetooth, NFC, etc. The function may be secured by one or more of a password, digital fingerprint (such as a cookie or token) required on a modifying device, or by software requiring physical identification or verification of a user's identity by a user's personal electronic device (e.g., an app on a smartphone or other mobile device) before allowing modification of the function's software. This method of allowing some degree of programming or reprogramming the gate keeping module would allow for beneficial security protocols and updates, not necessarily for "intrusion prevention" but also for "transit" related security and customization. For example, a gate keeping module might be programmed to connect to another cloud computing platform other than one designated and hardcoded by the original manufacturer (e.g., Alexa™/Echo® may be reconfigured to connect to a Microsoft® platform by identifying a new IP addresses to which audible speech is permitted to be sent).

The gate keeping module may also be programmed and/or updated regularly by predetermined software, such as (but not limited to) a smartphone "app." In some embodiments, the device may be reprogrammed to use a VPN (Virtual Private Network) to relay the audible speech or gesture data to a server processor in the cloud. Alternatively, the device may be programmed to select from among a set of possible IP addresses, based on load-balancing or traffic-balancing considerations. A security feature may be added to have a two-point verification of encryption and/or malicious code passing as or through as audible data or gesture data uploaded.

Alternatively, a gate keeping module or processor may be programmed to forward received data to a local preprocessor. For example, received speech or gesture data could be forwarded to a local server **207** on the same wireless network as the device, or to a smartphone app, instead of to a remote server. In this way, information privacy and security can be ensured, and may also allow operation in some capacity even if the wireless network's connection to the greater internet is lost.

In another embodiment, a digital personal assistant may comprise a plurality of processors where the wireless communication component is limited to receive externally-originated inbound data and transmitting the inbound signal to a first processor only, the first processor being unable to transmit signal to the wireless component and only able to transmit signals to a display or speaker. Further, the wireless communication component may be limited to only receive outbound (received input) data from a second processor (received from one or more receivers) to transmit to a remote computing device; thus, rendering the receivers unable to be overridden by digital intruders.

Purely local processing of speech and gesture data may enable more secure applications related to security and

privacy. For example, if a home security system is controlled by speech or gestures, local processing may be preferred to remote processing that might allow a remote hacker to disable the house's security. A private and secure "walkie talkie" or other communication function may be added to operate off the local network and allow communication between multiple devices on the local network.

In some embodiments, the preprocessor **201** may be programmed to allow continuous transmission of all audio or video data received by the receiver by default, and only enforce restrictions on the data transmission in response to particular user inputs.

In another embodiment of the preprocessor(s) and its function, the preprocessor **201** may contain a localized Natural Language Processing (NLP) programming embedded in a non-transient memory tasked with pre-parsing continuous strings of received input by the user into an individually most-comprehensible sub-string.

For example, a user may prefer to make a series of commands without pause, the always listening device and/or its cloud computing unit might not be able to decipher a complex series of commands such as: "Play classical music on Pandora® set volume to four stop playing in one hour turn off bedroom lights set alarm seven A.M." However, utilizing the preprocessor **201** comprised of a pre-parsing NLP may allow analysis and transformation of the string into five individual commands before transmitting to the cloud computing unit:

"Play classical music on Pandora®,"

"Set volume to 4,"

"Stop playing in 1 hour,"

"Turn off bedroom lights," and

"Set alarm 7 A.M."

In response, the cloud computing server **203** may be able to respond respectively in sequence, as normal to its original function and/or continuously:

"Playing classical music on Pandora®."

(Silently actualize change in setting—i.e., Amazon Alexa™ does not provide audible feedback on changing volume)

"I will stop playing in 1 hour."

(Actualize the change in setting) "Okay."

"Alarm set for 7 A.M."

The pre-parsing of a string may be accomplished at the remote cloud computing unit **203** as well; or in any other configuration where it is accomplished before being introduced to the actual NLP or ASR (Automated Speech Recognition).

In another embodiment, the gate keeping module and/or its associated hardware may be independent, in a device separate from an "always listening" electronic device.

The gate keeping module **201** may comprise Natural Language Processing (NLP) to locally process input from a user.

In FIG. 6, the arrows depict directional flow of data transmissions, both wireless and wired, within the electronic device. As shown, the preprocessor and/or gate keeping module **201** is restricted to single-direction, outbound transmissions. The gatekeeping function **201** may be embedded in either a single-directional processor (i.e., a device or component having only input from the receiver and having only output to a downstream processor) or a bi-directional processor, so long as the gate keeping module is restricted to only a single-directional, outbound transmission. The single-directional feature, if not enforced by the data input/output ports themselves, may be enforced by, for example,

software stored in a read-only memory (ROM) and executed by a processor in the gate keeping module **201**.

In the illustrated embodiment, the preprocessor and/or gate keeping module **201** relays and protects speech data received by the receiver **200** before transmitting it to the primary processor and/or wireless input/output communications component **202**, which transmits the speech data to the cloud or other remote server **203**.

The primary processor **202** may encrypt the speech data or otherwise ensure security of the data transmission channel prior to transmission. Alternatively, the encryption may occur at the preprocessor level **201**; or both, for a multilayer encryption feature.

The system/component/method may be integrated into various types of electronic devices, such as (but not limited to): mobile phones, tablets, laptops, computers, smart watches, televisions, lighting, media players (e.g., a DVD player, a Blu-Ray player, iPod, etc.). A person of ordinary skill in the art would be able to apply this invention to a number of other scenarios, applications and methods of integrating an embodiment that would provide added value to the user, device maker, content (and/or advertising) delivery provider, or a combination thereof.

In some embodiments, an integrated gate keeping module without external data inputs can be used to protect video or other types of data feeds that a user might find sensitive and wish to keep secure and private.

In some embodiments, the device may include one or more light emitting diodes (LEDs), whose lighted/unlighted status or color allows the user to visually ascertain whether the always listening mode is on or off. Other visual or auditory effects may be used to indicate the always listening status, including, by way of example, a symbol, icon, or flashing icon on a screen or display of a device, an occasional beep or prerecorded sound to remind the user that the always listening mode is engaged, or any other way of alerting or reminding a user via that user's sensory input that recording is occurring.

While a speaker for device output to the user is preferred, in some embodiments, the device might not include a speaker for output, which may instead be provided by one or more of lights, vibrations, a video screen, or other visual indicator. Alternatively, the device may be both completely silent and unchanging in visual appearance, exclusively performing data transmissions and updates in the background without feedback to the user.

In some embodiments, the device may incorporate one or more sensors in addition to a microphone or camera, including (but not limited to) a vibration sensor (such as a seismograph), global positioning system (GPS), accelerometer or gyroscope for determining orientation, thermometer, humidity sensor, etc. The additional sensor(s) may be used to determine possible user intent even without an utterance or gesture, such as, for example, detecting the vibration of a fallen user, an unsafe temperature in a living area, or other emergency situations.

The "always listening" mode, where and when the capability is enabled by the user, the device and/or its offsite functions (e.g., cloud computing, logic, Natural Language Processor, or artificial intelligence) may listen/receive/record, process, record useful data and ascertaining appropriate times to respond and/or provide assistive details. For example, if a user elects to have "always listening" on, the device may be able to provide a number of useful services:

If a user schedules an appointment (with another user physically present or via telephone), the device may add the appointment to a software calendar associated with the user.

If a user conversationally inquires to a second user, “Who is that actor?” while watching a movie, the device may attempt to determine the movie and scene being watched, check a data repository of information about the movie, and respond with a best guess/estimate of the identity of the actor.

If a user speaks to a secondary user or household member to “remind them” of an upcoming event or assigned task, the device may respond appropriately by adding a reminder regarding the event or task to their digital or cloud notes, or adding a mobile phone alert for that secondary user.

If a user converses with another user or household member on where to go or what to have for dinner, the device may respond with suggestions of nearby restaurants that have paid to advertise, have promotions, etc.;

If a user asks how far a suggested restaurant is, the device may determine the user’s location (via a GPS unit in the device, address from the user’s profile, or another means of estimation such as cell tower triangulation) and respond with approximate distance from the user to the restaurant.

If a user says the restaurant is too far (to the device or to another user), the device may inquire if there is a time or distance parameter the user would like to stay within.

If users discuss and pause for a period of time after the last time or distance conversation concluded, the device may make an alternative suggestion within any provided parameters or parameters received or determined from the users’ discussion.

If a user(s) discuss and pause for a period of time after a type of food the user(s) would like, the device may make a suggestion within the parameters of the user(s) discussion or instructions.

If a user(s) discuss and pause after a period of time a movie he/she/they would like to see, the device may provide information on show times, locations, distance or a combination of.

If a user conversationally describes the environmental conditions as being uncomfortable, the device may connect to one or more thermostats, cooling systems, heating systems, fans, or other external environment control devices, transmitting an electronic command to the apparatus’ to modify the settings to a more desirable level. Similarly, if a user were to say, “It’s too bright in here,” the device might transmit a signal to motorized curtains to close or close slightly, or transmit a signal to a dimmer switch to dim the lights to a lower setting.

If user were to fall and/or emit a cry of pain or discomfort, the device may inquire as to the well-being of the user, contact a third party such as a family member or friend, or contact an emergency service to direct medical aid to the user. If a user is unresponsive after a fall or injury, the device (if permitted by the user in advance within the systems’ settings) may override and access additional video feeds to ascertain the state of the user, and/or forward the video feeds to the third party or emergency service.

If a user describes symptoms of illness, the device may suggest methods of treatment, automatically update a user’s medical record for consultation during a future

medical appointment, and/or interject, recommending seeking immediate medical attention.

The user may preprogram the system and/or device to recognize specific “safety” and “emergency-send help” code words or gestures to enhance the security of the user. For example, a code word such as “blue elephant” may be used as an “all-safe” word/phrase (and may be verified with Voice ID) upon entry of a dwelling equipped with an armed security system. Alternatively, a user may configure a specific phrase or gesture to indicate “User under duress—send help!” If an intruder subsequently coerces a user to deactivate a security system, uttering the phrase may cause the security system to enter a false disarm state while alerting the authorities. A third phrase or gesture may be used to immediately trigger the security system, as a “panic” command while the user attempts to escape or hide.

If a user is unable or unwilling to speak (e.g., the user is deaf, is mute, is eating, is in a loud environment and does not want to shout over the sound, etc.), the user may still enjoy the safety and helpfulness of a personal assistant device through custom or language based gesture communications with the device.

Embodiments controlling the continuous transmission of video content may be useful in an industrial/commercial scenario—for example, where theft may occur or safety/liability is a concern. In a warehouse implementing an always listening or watching system, the data feed (audio, video or both) may be continuously processed, deciphered and/or analyzed to formulate, retrieve, compute, access, assess or a combination thereof, for a best suitable response to the data such as providing visual feedback through a display or audio feedback through speakers to aid workers, enhance safety or detect unsafe behavior, or alert store or official authorities to break-ins, shoplifting, or theft. Additionally, the system may save received data regarding infractions for training purposes.

Alternatively, the received data used in monitoring infractions may be further used for determining non-moving violations such as illegal parking or exceeding a time limit for parking meters, etc.

A response by the system and method may be passive or active. For example, a passive response would be, in response to an overheard conversation, making an appointment to the user(s) calendar(s); an active response would be providing restaurant suggestions to an overheard conversation of where to go for dinner or order delivery.

Some embodiments may contain one or more displays to provide video on demand or assistance delivery, such as, for example, visual how-to’s, advertisements, promotions, coupons or similar.

In some embodiments, other data about a user that comes up in conversation, such as a user’s favorite color, favorable or unfavorable opinion on a topic, or other commentary that comes up in conversation, debates, or arguments may be recognized by a processor as not a query or a request to which assistance may be provided. However, the information may nonetheless be archived in a database to allow for better assistance in the future, for example, in response to a request to purchase an item online, purchasing the item in the user’s favorite color, or suggesting a book or TV show to watch based on previous commentary by the user. The system may, in some embodiments, learn and store one or more of a user’s age, gender, household demographic, products owned, and other information relevant to a user’s tastes. Data about users in a household may be anonymized



to prevent identification and respect privacy of the users in the case of data breach or other unauthorized access to the data stores.

Audible data collected from a user through an always listening device and the response of assistance might not be restricted to households only. For example, in an industrial or commercial application, audible data collected through an always listening device, may be deciphered and analyzed. If one employee asks another, "Do we have a certain item in stock?", the system may interject in the conversation and audibly respond, "Yes, 123 units are available in the New Jersey warehouse and 234 units are available in the New York warehouse" or "No, but there is an expected shipment due in July 7th." By building up a model and learning the context of conversations held in a particular device's location, responses to ordinary questions can be accurately provided.

FIG. 7 depicts, in simplified form, another embodiment of an "always-listening" device.

In some embodiments, the preprocessor and/or gate keeping module 301 can be integrated together with the primary processor 306 and wireless I/O communication module 302 so long as the preprocessor and/or gate keeping module 301 is restricted to have only single-direction, outbound transmission capability. In this embodiment, the primary processor wireless transmitter/receiver 302 transmits the speech data to cloud or other remote server device 303 and receives a response in turn that may be relayed in the form of audible feedback through one or more speaker(s) 304. Power from power source 305 may still, in some embodiments, only pass to microphone 300 via the integrated gatekeeping module/processor/communication module, while in other embodiments, the microphone 300 may have a direct power link that is not severable.

FIG. 8 depicts, in simplified form, an alternative embodiment of an "always listening"-capable electronic device.

In some embodiments, a first preprocessor and/or gate keeping module 401 and a second preprocessor and/or gate keeping module 402 may each be passively "always listening" for its own unique wake up phrase. For example, the preprocessor and/or gate keeping module 401 may have the wake up phrase be the word "Alex," whereas the primary processor wake up word may be "Alexa™" (per usual for an Amazon Echo®).

In other embodiments, the wake up phrase may be a particular word, a particular series of words, a hand gesture, a facial gesture, a bodily movement, a tone, a whistle, a pattern of sounds, a pattern of movements, a pattern of taps, or any combination of the above. The wake up phrase may be preconfigured to the device, or may be set by a user to the user's taste.

In one example, if a user says "Alex," followed by a single string such as "allow always listening," the system may respond by allowing continuous listening and transmitting of speech data. Similarly, "Alex" followed by "stop listening" may discontinue listening and transmitting of speech data.

After fulfilling the "stop listening" command, the system may return to a passive mode. The preprocessor and/or gate keeping module may also be integrated or kept independent of a second or primary processor. Further, the wireless I/O or transmitter/receiver may be integrated or kept independent of a processor.

In some embodiments, the first gate keeping module 401 may have control over or communicate with the second gate keeping module 402. For example, if gatekeeping module 401 is added to an existing system that is not fully trusted,

a user may be able to use a first wake up phrase to control whether the existing system is able to listen at all, and prevent the existing system from reporting recorded data to a central server.

FIG. 9 depicts an illustrative flow chart of an improved method of processing a second, distinct wake up word.

For example, in the case of an Amazon Echo®, a second wake up word, such as "Alex," could be used in addition to the default "Alexa™." While listening (Step 900), if the first default wake up word is used (Step 901), the standard device behavior is followed.

The next audible phrase after the received wake up word is received (Step 903) and transmitted to the cloud or other remote server (Step 904) for natural language processing (NLP) and a determination by the remote server of the best direct or indirect response to each verbal input or gesture by the user (Step 905). If the proper response is information, speech, music, etc., it is output through the device's speaker(s); if the proper response is an action (e.g., downloading a file, communicating with another device's API, etc.), that action is performed (Step 906). The device then continues to listen again for the first wake up word or gesture (Step 900), process any additional verbal input, and deliver any subsequent assistance to the user.

If, instead, the second wake up word is received (Step 902), the "always listening" mode is engaged (Step 907). The always listening cycle (Steps 908-911) mirrors the conventional series of steps (receive user input, transmit to cloud, determine best response, and effect the response), but in a constant cycle, or even a constant series of multiple concurrent cycles (i.e., receiving a second user input and transmitting it even before the server responds based on the first input), instead of returning to waiting, dormant, for a wake up word. Multiple, asynchronous series of the steps (Steps 908-911) may be in processing simultaneously, and responses may be prioritized in real time based on perceived user need. For example, if a user were to say "I wonder where the nearest Mexican restaurant is. Actually . . . I think I want Chinese instead," the system might have begun processing the request to determine a Mexican restaurant location, but then received the second statement before making the determination. In response to receiving the second statement, the Chinese restaurant query may be processed and output, in lieu of the Mexican restaurant.

Responses may be prioritized based on any one or more of perceived indicators of urgency (e.g., tone, volume, vocabulary used to indicate importance or danger), emergency or safety concerns, content and subject relevancy, determination of the user's current intent or interest, determining that a user no longer needs assistance, ranking users in a set of users (such as responding first to a device's owner, or to adults over children), subject relevancy, or highest bidder of advertisement. Thus, after the "always listening" mode is engaged by receipt of the second wake up word, the system can become a passive or active participant in the conversation, interjecting to query or provide information, or passively updating calendars or other data stores in response to what is overheard.

FIG. 10 depicts an alternative flowchart of program logic for additional features of an always-listening device.

As in FIG. 9, the system listens for input (Step 1000); if a first wake-up phrase is used, the normal, single-query process may be followed (Steps 903-906), and if a second wake-up phrase is used, in conjunction with a command to engage always-listening mode, the always-listening mode with continuous response to input may be triggered (Steps 908-911). This feature may prevent users from accidentally

13

triggering always-listening mode merely by saying the second wake-up phrase, without confirming their intent, and unwittingly having the device be listening and/or performing actions in the background.

In addition, the always-listening mode should always be listening for an instruction of the user to end the always-listening mode (Step 1005), in which case the system may return to its normal function of listening (Step 1000) in order to enter either a single-query cycle or a multi-query, always listening cycle.

In some embodiments, the system may listen for an additional command to disable one or more parts of the system functionality (Step 1002). The functionality involved may include disabling a microphone or other receiver of the device, disabling a gatekeeping module (to allow data or power to be transferred through the gatekeeping module unhindered), disabling one processor of a plurality of processors, or disabling a communication module for communicating with the cloud or with local devices. The system can then listen for a command (Step 1003) to re-enable the disabled functionality, which, if received (Step 1004) returns the system to its previous function.

The disable functionality/enable functionality cycle (Steps 1002-1004) may instead be included within the flow of either the single-query cycle (for example, between Steps 903 and 904) or multi-query always-listening mode (for example, between Steps 908 and 909).

For example, while in always-listening mode, a command could be issued to turn off one microphone or other receiver used by a multi-receiver device, while leaving another microphone or receiver functional, if, for example, a user does not trust an underlying digital personal assistant receiving data from the first receiver, but does trust a gatekeeping module added to the digital personal assistant and controlled by input from a second receiver.

The descriptions of the various embodiments of the present invention have been presented for purposes of illustration, but are not intended to be exhaustive or limited to the embodiments disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the described embodiments. The terminology used herein was chosen to best explain the principles of the embodiments, the practical application or technical improvement over technologies found in the marketplace, or to enable others of ordinary skill in the art to understand the embodiments disclosed herein.

The present invention may be a system, a method, and/or a computer program product at any possible technical detail level of integration. The computer program product may include a computer readable storage medium (or media) having computer readable program instructions thereon for causing a processor to carry out aspects of the present invention.

The computer readable storage medium can be a tangible device that can retain and store instructions for use by an instruction execution device. The computer readable storage medium may be, for example, but is not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. A non-exhaustive list of more specific examples of the computer readable storage medium includes the following: a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a static random access memory

14

(SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disk (DVD), a memory stick, a floppy disk, a mechanically encoded device such as punch-cards or raised structures in a groove having instructions recorded thereon, and any suitable combination of the foregoing. A computer readable storage medium, as used herein, is not to be construed as being transitory signals per se, such as radio waves or other freely propagating electromagnetic waves, electromagnetic waves propagating through a waveguide or other transmission media (e.g., light pulses passing through a fiber-optic cable), or electrical signals transmitted through a wire.

Computer readable program instructions described herein can be downloaded to respective computing/processing devices from a computer readable storage medium or to an external computer or external storage device via a network, for example, the Internet, a local area network, a wide area network and/or a wireless network. The network may comprise copper transmission cables, optical transmission fibers, wireless transmission, routers, firewalls, switches, gateway computers and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer readable program instructions from the network and forwards the computer readable program instructions for storage in a computer readable storage medium within the respective computing/processing device.

Computer readable program instructions for carrying out operations of the present invention may be assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, configuration data for integrated circuitry, or either source code or object code written in any combination of one or more programming languages, including an object oriented programming language such as Smalltalk, C++, or the like, and procedural programming languages, such as the "C" programming language or similar programming languages. The computer readable program instructions may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider). In some embodiments, electronic circuitry including, for example, programmable logic circuitry, field-programmable gate arrays (FPGA), or programmable logic arrays (PLA) may execute the computer readable program instructions by utilizing state information of the computer readable program instructions to personalize the electronic circuitry, in order to perform aspects of the present invention.

Aspects of the present invention are described herein with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems), and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer readable program instructions.

These computer readable program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the com-

15

puter or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer readable program instructions may also be stored in a computer readable storage medium that can direct a computer, a programmable data processing apparatus, and/or other devices to function in a particular manner, such that the computer readable storage medium having instructions stored therein comprises an article of manufacture including instructions which implement aspects of the function/act specified in the flowchart and/or block diagram block or blocks.

The computer readable program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other device to cause a series of operational steps to be performed on the computer, other programmable apparatus or other device to produce a computer implemented process, such that the instructions which execute on the computer, other programmable apparatus, or other device implement the functions/acts specified in the flowchart and/or block diagram block or blocks.

The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods, and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). In some alternative implementations, the functions noted in the blocks may occur out of the order noted in the Figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts or carry out combinations of special purpose hardware and computer instructions.

What is claimed:

1. An always-listening-capable computing device, comprising:

- a first receiver configured to receive input from a user comprising an utterance or a gesture;
- a communication module configured to transmit data to and receive data from a remote server;
- a gate-keeping module implemented by a processor, wherein all data received by the communications module based on data from the first receiver passes through the gate-keeping module while a gatekeeping function is disabled, wherein no data based on data from the first receiver passes through the communications module while the gatekeeping function is enabled, wherein all data input to the gate-keeping module is received via an exclusive input lead from the first receiver, wherein all data output from the gate-keeping module is transmitted via an exclusive output lead to a component other than the first receiver; and

non-transient memory storing instructions that, when executed by a processor, causes a processor to:

- determine that the input received by the first receiver comprises a first wake up phrase; and

16

in response to determining that the input comprises the first wake up phrase, automatically perform at least one of:

- transmit a representation of user input subsequent to the first wake up phrase;
- activate an always-receiving mode to transmit a stream of user input captured subsequent to the first wake up phrase;
- deactivate the always-receiving mode to prevent transmission of user input received subsequent to the first wake up phrase, unless also preceded by a second wake up phrase;
- or enable the gate-keeping module to prevent all transmission of data external to the always-listening-capable computing device.

2. The always-listening-capable computing device of claim 1, wherein the first receiver comprises one or more power input leads and one or more data output leads to the gate-keeping module, and wherein the first receiver does not comprise any input leads from or output leads to other components of the always-listening-capable computing device.

3. The always-listening-capable computing device of claim 1, wherein the gate-keeping module comprises one or more power input leads.

4. The always-listening-capable computing device of claim 1, further comprising two or more gate-keeping modules, wherein a first gate-keeping module of the two or more gate-keeping modules is enabled or disabled by a second gate-keeping module of the two or more gate-keeping modules.

5. The always-listening-capable computing device of claim 1, further comprising a second receiver, wherein the gate-keeping module activates the second receiver, in response to the first receiver receiving the first wake up phrase, wherein a processor of the one or more processors transmits data in response to input received by the one or more receivers comprising a second wake up phrase different from the first wake up phrase.

6. The always-listening-capable computing device of claim 1, wherein the gate-keeping module is configured to prevent the communication module from transmitting data external to the always-listening-capable computing device by preventing a power source from powering the communication module.

7. The always-listening-capable computing device of claim 1, wherein the always-receiving mode comprises:

- analyzing, by the remote server, the transmitted data to determine a response comprising at least one of:
- actuating a determined possible user desire;
- soliciting additional information from the user to enable a response;
- archiving non-identifying data from the user; and
- providing responsive information to the user.

8. The always-listening-capable computing device of claim 1, wherein the non-transient memory stores instructions that, when executed by a processor, further causes the processor to:

- enable, for a first interval of time, the communication module to receive data;
- disable, for a second interval of time shorter than the first interval of time, the communication module from receiving all incoming wireless data; and
- re-enable the communication module to receive incoming wireless data.

9. The always-listening-capable computing device of claim 1, comprising two or more receivers, wherein the two

17

or more receivers comprise a microphone and one of a camera and a sensor of movement.

**10.** A computer-implemented method for providing always listening assistive response, comprising:

receiving, via a receiver, user input comprising one or more utterances or gestures;

passing data based on received user input to a communications module via a gatekeeping module implemented by a processor, such that all data received by the communications module based on data from the receiver passes through the gate-keeping module while a gatekeeping function is disabled, wherein no data based on data from the receiver passes through the communications module while the gatekeeping function is enabled, wherein all data input to the gate-keeping module is received via an exclusive input lead from the receiver, wherein all data output from the gate-keeping module is transmitted via an exclusive output lead to a component other than the receiver;

determining that the user input comprises a first wake up phrase; and

in response to determining that the user input comprises a first wake up phrase, automatically performing at least one of: transmitting a representation of user input subsequent to the first wake up phrase;

activating an always-receiving mode to transmit a stream of user input captured subsequent to the first wake up phrase;

deactivating the always-receiving mode to prevent transmission of user input received subsequent to the first wake up phrase, unless also preceded by a second wake up phrase;

or enabling the gate-keeping module to prevent all transmission of data by the communications module.

18

**11.** The computer-implemented method of claim **10**, wherein the user input comprising one or more utterances or gestures comprises user input from a plurality of users.

**12.** The computer-implemented method of claim **10**, further comprising:

after transmitting user input captured subsequent to the first wake up phrase, displaying to a user one of: a portion of a web page; a fact; a calculation; marketing assistance; a subject-relevant, time-applicable advertisement; music; a data stream comprising real-time communications from a third party, or notification that an action has been performed.

**13.** The computer-implemented method of claim **10**, wherein the device tracks a conversation state and prompts a user with one or more clarifying queries to determine more appropriate information to satisfy a user desire.

**14.** The computer-implemented method of claim **10**, further comprising silently causing a remote data store associated with the user to be updated.

**15.** The computer-implemented method of claim **10**, wherein the gate-keeping module is incorporated into a digital personal assistant system, wherein a receiver of the digital personal assistant system is disabled until a wake up utterance or gesture is processed by the gate-keeping module, the wake up phrase being distinct from a second wake up phrase preconfigured to be used to activate the digital personal assistant system.

**16.** The computer-implemented method of claim **15**, wherein the digital personal assistant system is disabled via depowering a receiver that is configured to receive user input.

**17.** The computer-implemented method of claim **15**, wherein the digital personal assistant system is disabled via the gate-keeping module preventing transmission of data to the digital personal assistant system.

\* \* \* \* \*