August 12, 2019

PANIC WORD

Tajinder Gadh

Marc Stogaitis

PANIC WORD

ABSTRACT

A computing device (e.g., a mobile phone, camera, tablet computer, etc.) may include an integrated display device (e.g., a presence-sensitive screen) at which a user interface is presented. Additionally, a computing device may include a microphone that generates audio data and the capability to process the audio data. For instance, the computing device may process the audio data to identify one or more commands or requests spoken by a user of the computing device, and perform various actions associated with those commands or requests. In some situations, a user may desire for the computing device to perform one or more actions that would typically be performed via interacting with the displayed user interface without having to touch the phone. For instance, in an emergency situation, it may be desirable for the user to cause the computing device to contact emergency services (e.g., call 911 or a local variant of such number) without having to physically interact with the device. The user may utter a specific word or group of words which the computing device recognizes to cause the computing device to contact emergency services.

DESCRIPTION

Computing devices may include functionality to detect and analyze audio data to recognize the utterance of a specific word or group of words that cause the device to execute an action or set of actions. The set of specific words the device recognizes which cause it to execute such actions may be referred to as "hot words." To illustrate, a computing device may analyze audio data to determine when a user says the hot words, "Ok computer." Upon recognizing that the user uttered this hot word, the computing device may process any commands or requests the user utters after the hot word. To further illustrate the previous example, if the user says, "Ok computer, start a five minute timer," the computing device may cause the start of a five minute timer on the device.

As demonstrated by the example, a purely verbal interaction with the computing device allows a user to cause certain actions without having to physically interact with the device. This may benefit a user who is unable to use their hands in a given situation, such as while holding a hot pan, driving a vehicle, or experiencing an emergency. However, existing computing devices typically do not allow for hot words to be commands themselves, i.e. cause the computing device to execute certain actions. Instead, the hot words only alert the computing device to process subsequent audio data for commands. To permit faster computing action response, it may be desirable for a computing device to execute certain actions using the utterance of hot words only. Such applications may be beneficial in emergency situations when a user may be unable to physically interact with the device, or when prolonged or unnatural interaction with the device or displayed user interface may slow the user's escape from harm or alert a perpetrator that emergency services have been contacted.

The example computing device shown in Figure 1 is configured to execute certain actions based on the utterance of specific hot words. For example, in response to determining that a user says, "Help! Help! Help!" the computing device may initiate an action to contact emergency services. The example computing device of Figure 1 includes a user interface module, one or more input and output devices, one or more microphones, a speech recognition module, a machine learning model, and/or one or more other modules. Example computing devices include, but are not limited to, a smartphone, tablet computer, laptop computer, computerized watch, computerized eyewear, personal computer, smart television, personal digital assistant, automobile navigation and/or entertainment system, vehicle (e.g., automobile, aircraft) and/or cockpit display, a home or other smart appliance and/or related device (e.g., interconnectable appliance/device via Internet of Things), or any other type of wearable, non-wearable, mobile, or non-mobile computing device.

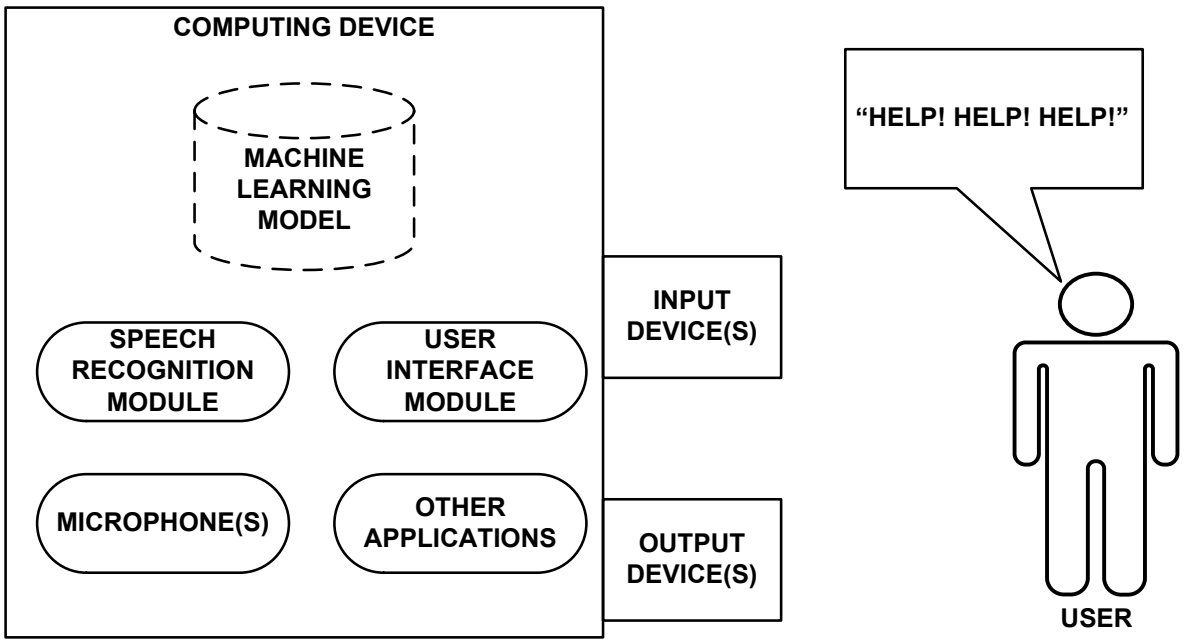

Figure 1

The example computing device of Figure 1 may include a user interface module. The user interface module may include a user interface device (UID) to function as an input and/or output device for the computing device. The UID may function as an input device using a presence-sensitive input screen, such as a resistive touchscreen, a surface acoustic wave touchscreen, a capacitive touchscreen, a projective capacitance touchscreen, a pressure sensitive screen, an acoustic pulse recognition touchscreen, or another presence-sensitive display technology. The UID may function as an output (e.g., display) device using any one or more display devices, such as a liquid crystal display (LCD), dot matrix display, light emitting diode (LED) display, organic light-emitting diode (OLED) display, e-ink, or similar monochrome or color display capable of outputting visible information to a user. The user interface module controls the user interface device including determining what the user interface device presents and what information is exchanged between the user interface device and other applications or components of the computing device.

As shown in Figure 1 above, the computing device may include one or more input devices and one or more output devices. For instance, the input devices may include wired or wireless headphones which may include a microphone for detecting audio data, a mouse, a keyboard, a voice responsive system, a camera, a fingerprint sensor, or any other type of device for detecting input from a human or machine. In some cases, the input device may include one or more location sensors (e.g. GPS components, Wi-Fi components, cellular components), one or more temperature sensors, one or more movement sensors (e.g., accelerometers, gyroscopes), one or more pressure sensors (e.g., barometer), one or more ambient light sensors, and/or one or more other sensors (e.g., camera, infrared proximity sensor, hygrometer, and the like). Other sensors may include a heart rate sensor, magnetometer, glucose sensor, hygrometer sensor, olfactory sensor, compass sensor, step counter sensor, to name a few other non-limiting examples.

The computing device may also include one or more output devices, such as one or more speakers or display screens, including a presence-sensitive screen (e.g., touchscreen), flash of a camera, auxiliary port, or any other type of device for generating output to a human or machine. In some cases, the input devices and/or output devices may include one or more other types of wearable, non-wearable, mobile, or non-mobile computing devices that are also used by the user. One or more of the input and/or output devices may be external to and communicatively coupled (e.g., via a wired or wireless connection) with the computing device.

The computing device may also include one or more microphones, configured to generate audio data from the ambient surroundings. The microphone may be, include, or be used in combination with a dynamic, electret condenser, micro-electrical mechanical system (MEMS), fiber optic, and/or piezoelectric type microphone. The microphone may be used to allow a user to make phone calls, record audio data of the device surroundings, if for example, recording the audio

data of a meeting, or the microphone may be used to identify one or more commands or requests dictated by the user to the computing device.

The computing device may be configured to manage inputs received by the one or more microphones to be processed by the speech recognition module. This may allow users to verbally interact with the computing device. The speech recognition module may therefore function as an interactive assistant. The interactive assistant may be a voice-assistant that receives audible user commands, processes the commands with the speech recognition module, and executes corresponding actions, such as providing audible responses to user queries and/or executing certain actions such as displaying information with the user interface module.

When a user provides audible input to the interactive assistant (e.g., commands, questions, queries), the interactive assistant may use the speech recognition module to process such audible input. The computing device may contain a machine learning model to process the audio data from the one or microphones. The machine learning model may be stored on the device, or it may be stored on an external computing device, with which the computing device is communicatively coupled. The machine learning model may be configured to recognize a specific user's voice, such as the primary user of the device, while ignoring other audio inputs. For example, the computing device may respond when the primary user says "Ok, computer," but ignore the command when a secondary user utters the same hot word "Ok, computer." This may allow for the computing device to only execute actions the primary user of the device desires to be performed.

As described herein, the speech recognition module and user interface module are configured to utilize and/or process information received from the input devices only after receiving explicit authorization from the user to do so. The speech recognition model typically

uses hot words to authorize the computing device to execute a command. After receiving such authorization, the computing device may execute the desired command.

The example computing device of Figure 1, using the one or more microphones and the speech recognition module, may detect audio data from ambient surroundings. Such audio data typically includes user-dictated speech. When the user utters a specified word, or group of words, presently described as hot words, the computing device may execute an action or series of actions. Instead of waiting for additional commands from the user, as typical computing devices currently do, after recognizing the utterance of hot words, the computing device may immediately execute an action or set of actions. For example, as shown in Figure 1, a computing device may be configured to execute an action immediately after a user utters "Help! Help! Help!" without waiting for additional verbal commands. This allows for the computing device to more quickly respond to user commands. Such an action may be particularly beneficial when the user is in an emergency situation.

Upon recognizing the utterance of hot words, the computing device may contact emergency services on behalf of the user without requiring the user to physically interact with the device. Emergency services contacted by the computing device may include law enforcement officers and officials, fire and rescue services, emergency medical services, and the like. The computing device may also be configured to contact a list of emergency contacts stored on the computing device. For example, a user may store emergency contacts on the device. Such contacts may include family members, close friends, or other groups or individuals a user may desire to be contacted in emergency situations.

The computing device may contact emergency services or emergency contacts through text message, audio message, phone call, and/or other notification methods. For example, after a user

says “Help! Help! Help!” the computing device may place an automated call to 911 or other similar service on behalf of the user. The device may also send a text message or other notification to the user’s emergency contacts to provide them with notice of the user’s circumstances. These actions executed by the computing device may allow for the user to quickly notify emergency services and other individuals without having to physically interact with the device, or require the user to utter hot words plus an additional command. Uttering an additional command after the hot word may not be feasible in a given situation, or may alert a perpetrator that the user is attempting to contact emergency services.

The computing device may also be configured to capture audio data after the hot words are uttered by a user. The recorded audio may then be sent to emergency services or emergency contacts. For example, once the user dictates the hot words, the computing device may record a specified duration of audio data. Then, for example, if the computing devices places an automated call to emergency services, the computing device may include the audio data in the call. This may allow for emergency services to understand what the emergency is. For example, if the audio data records the user and another individual sounding like a struggle for personal property, emergency services may recognize the user is being robbed, and the operator should send law enforcement services. In another example, a user may state after uttering the hot words that they are trapped under a tractor on their farm and unable to move. This may alert emergency services the best course of action is to send emergency rescue and medical services to the user.

The computing device upon recognizing the utterance of hot words may also execute other actions to benefit the user. For example, the computing device may use a flash of a camera of the computing device to blink repeatedly. This may deter an attacking person or animal, notify those around the user of the user’s distress, or serve as a beacon for arriving emergency services or

contacts to locate the user. The computing device may also use the camera of the device to capture images or video of the surroundings of the computing device. For example, after the user utters "Help! Help! Help!" the device may use front and/or rear-facing cameras to take pictures and send them. This may provide an image of a third-person holding the user's device, a location of a user, if for example, they are trapped under a tree in the wilderness, or other useful information. The captured video or images may be provided to emergency services or emergency contacts to aid in the help of the user or to aid in locating the user's device, if for example, it is stolen.

The computing device may also be configured to send location or other relevant data to emergency services or emergency contacts. Many computing devices include GPS sensors which provide the device with a location of the computing device. Upon recognizing the hot words, the computing device may send location data to emergency services or emergency contacts on behalf of the user. Such an action would allow for emergency services to more readily locate and aid the user.

The computing device may also be configured to preserve user data after the hot words are uttered by the user. For example, after the user says "Help! Help! Help!" the device may encrypt the data of the device. The device may also preserve the data by deleting personal information, if, for example, a user is robbed of their device. The computing device may also place itself in a lockdown state where major permissions cannot be changed without the user verifying certain security information. Such actions by the computing device may protect the user's confidential, personal information stored on the device. This may protect a user so that the thief or later possessor of the computing device is unable to further harm the user by stealing their identity or otherwise using the data in an unsatisfactory way.

The user may configure the speech recognition module of the computing device to respond to a user-defined hot word, rather than a device-defined hot word. For example, the user may configure the device to execute certain actions when the user says "Help! Help! Help!" Alternatively, the user may configure the device to execute certain actions when the user says "I am in an emergency!" Allowing for a user-defined hot word to execute actions places the user in a better position to utter the hot word successfully in a panicked situation. If the user is unable to quickly remember the hot word when in such a situation, the speech recognition module is unable to help the user in an emergency situation. Allowing for a user-defined hot word may also allow the user to avoid alerting a perpetrator that emergency services have been contacted. For example, if the user programs the hot word to be "purple rainbows" a robber may not realize that the user was able to contact emergency services. This may help protect the user by allowing emergency services to reach the user before any harm is inflicted on the user.

The user may also configure the computing device to respond to their voice only. The machine learning model of the computing device may learn what the user's voice sounds like. For example, if the user is around a group of friends, and a friend says, "Help! Help! Help!" the machine learning model may recognize that this hot word did not come from the user, and not execute the action. The machine learning model may also be configured to, upon hearing the hot word from an individual not the user, execute the action as if it were the user, but include a message to emergency services that the individual in distress is not the primary user. This may provide information to emergency services enabling them to aid the proper individual in an emergency.

To ensure the computing device does not contact emergency services when not required, the computing device may incorporate validation methods into the action. For example, upon recognizing the hot word, the computing device may, through the user interface module, display a

time-decaying bar on the user interface to allow a user to abort the action. The abort command may require the user to validate their intention to abort by requiring a passcode or password, validating a user fingerprint, or validating intent via some other method. The computing device may also be configured to prompt the user to verbally abort the command after the user dictates the hot word. For example, if the user utters the hot word, through the speaker output device, the computing device may ask the user, "Are you sure you would like to contact emergency services?" If the user is either silent, or says "Yes," the computing device may then contact emergency services and/or emergency contacts. However, if the user replies, "No," then the computing device may abort the command. As discussed previously, the speech recognition module and machine learning module may be configured to recognize the user's voice. As such, the device may be configured so that another individual may not abort the action on behalf of the user.

The computing device may also be configured to learn and provide feedback to the user about the use of the hot word feature. For example, if the user utters the hot words multiple times over a period of time, and cancels the action each time, the computing device may recommend the user disable the feature. The computing device may also suggest the user opt-in if the user has not already done so to ensure the feature is available to the user when they may need the service most.

By configuring a computing device to execute certain actions with the use of hot words only, the computing device may help ensure the safety of the user. It may enable a user to quickly and efficiently contact emergency services without alerting a potential perpetrator. It may also benefit a user by allowing them to contact emergency services without having to physically interact with the device. The above examples are not limited to the described examples, as other hot words and means for executing actions with the dictation of only hot words may be utilized in similar ways to execute a variety of computing device actions.

It is noted that the techniques of this disclosure may be combined with any other suitable technique or combination of techniques.