June 2021

VOICE USER INTERFACE BASED PERMISSION GRANT SYSTEM FOR VEHICLES

Xin Li

Zheng Kou

Sriram Natarajan

Sukhyun Song

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

VOICE USER INTERFACE BASED PERMISSION GRANT SYSTEM FOR VEHICLES

ABSTRACT

A vehicle (e.g., automobile, motorcycle, a bus, a recreational vehicle (RV), a semi-trailer truck, a tractor or other type of farm equipment, a train, a plane, a helicopter, etc.) may include a so-called "head unit" that provides a voice user interface (VUI) by which to enable spoken human interaction with the head unit to respond to requests for permission (e.g., to access user personal data, to enable the usage of third-party services, etc.). For example, responsive to detecting that an action to be performed has not been granted permission, the head unit may produce (e.g., via one or more speakers) an audio prompt requesting the required permission. A user may answer the audio prompt with an audio input in the form of human speech, which the head unit may receive (e.g., via one or more microphones). The head unit may parse the audio input using speech recognition (e.g., a natural language understanding module) to identify a valid input (e.g., grant or deny permission to perform an action, request additional information, etc.) to which the audio input corresponds and, responsive to identifying a valid input, the head unit may perform the action (e.g., granting or denying permission to perform the action, providing additional information, etc.) associated with the valid input. In this way, the head unit may enable the user to control the granting of permissions via the VUI, which may be particularly beneficial in vehicle settings in which the user is operating the vehicle, as the hands-free, eyes-free user experience may reduce distractions to the user while operating the vehicle and thereby promote safety.

DESCRIPTION

FIG. 1 below is a conceptual diagram illustrating a system 10 including a head unit 100 of a vehicle (e.g., automobile, motorcycle, a bus, a recreational vehicle (RV), a semi-trailer truck, a tractor or other type of farm equipment, a train, a plane, a helicopter, etc.) and a computing system 120. As shown in FIG. 1, head unit 100 includes one or more processors 102, a display 104, one or more communication components 106 ("COMM components 106"), one or more speakers 108, one or more microphones 110, and one or more storage devices 112. As further shown in FIG. 1, computing system 120 includes one or more processors 122, one or more communication components 126 ("COMM components 126"), and one or more storage devices 132.
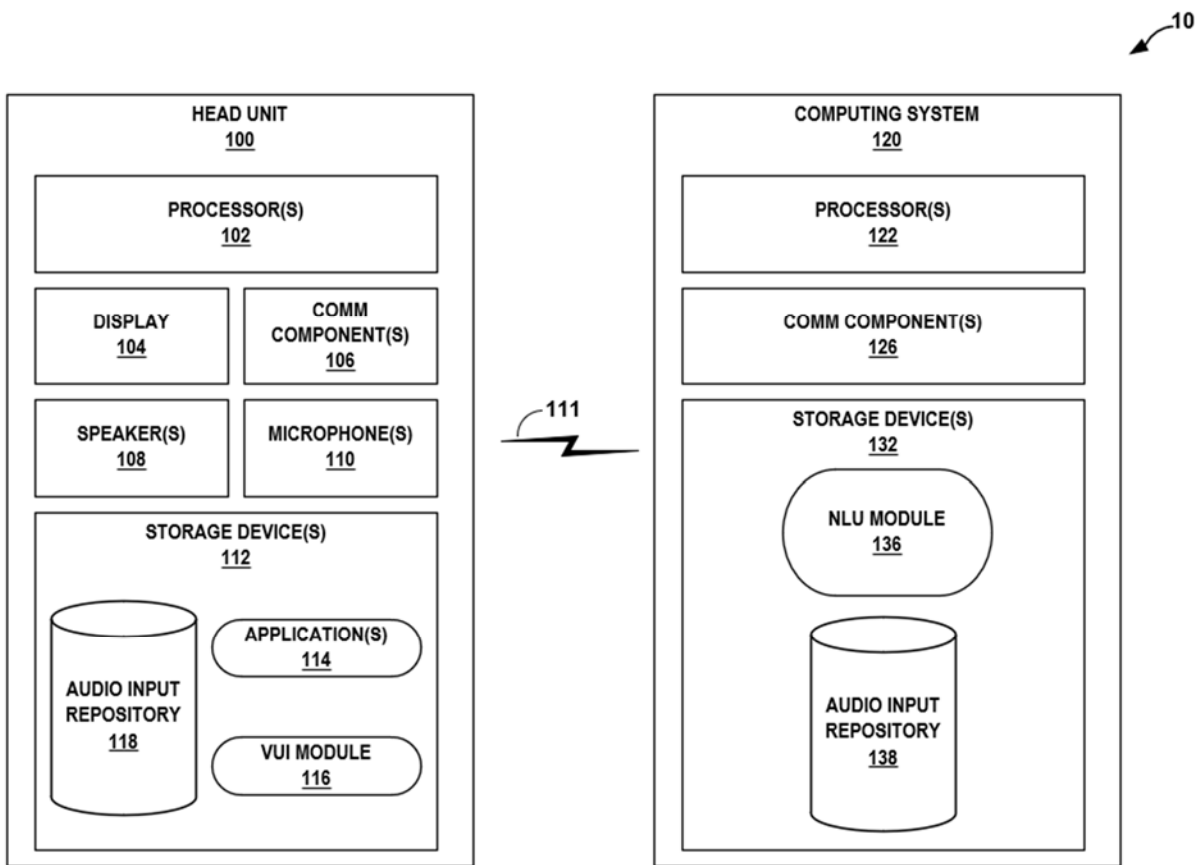

FIG. 1

Head unit 100 of system 10 may represent an integrated head unit that provides an interface (e.g., a voice user interface (VUI), a graphical user interface (GUI), etc.). In general, head unit 100 may control one or more vehicle systems, such as a heating, ventilation, and air conditioning (HVAC) system, a lighting system (for controlling interior and/or exterior lights), an infotainment system, a seating system (for controlling a position of a driver and/or passenger seat), etc. Head unit 100 may be included in a motorcycle, a bus, a RV, a semi-trailer truck, a tractor or other type of farm equipment, a train, a plane, a helicopter, a personal transport vehicle, or any other type of vehicle.

Processors 102 may implement functionality and/or execute instructions associated with head unit 100. Examples of processors 102 may include one or more of an application specific integrated circuit (ASIC), a field programmable gate array (FPGA), an application processor, a display controller, an auxiliary processor, a central processing unit (CPU), a graphics processing unit (GPU), one or more sensor hubs, and any other hardware configure to function as a processor, a processing unit, or a processing device. One or more applications 114 and a voice user interface module 116 ("VUI module 116") may be operable by processors 102 to perform various actions, operations, or functions of head unit 100.

Display 104 of head unit 100 may be a presence-sensitive display that functions as an input device and as an output device. For example, presence-sensitive display 104 may function as an input device using a presence-sensitive input component, such as a resistive touchscreen, a surface acoustic wave touchscreen, a pressure-sensitive screen, an acoustic pulse recognition touchscreen, or another presence-sensitive display technology. Additionally, presence-sensitive display 104 may function as an output (e.g., display) device using any of one or more display

components, such as a liquid crystal display (LCD), dot matrix display, light emitting diode (LED) display, active-matrix organic light-emitting diode (AMOLED) display, etc.

COMM components 106 of head unit 100 may include wireless communication devices capable of transmitting and/or receiving communication signals, such as a cellular radio, a 3G radio, a 4G radio, a 5G radio, a Bluetooth® radio (or any other PAN radio), an NFC radio, or a Wi-Fi™ radio (or any other wireless local area network (WLAN) radio). COMM components 106 may be configured to send and receive information via a network 111 (e.g., a local area network (LAN), wide area network (WAN), a global network, such as the Internet, etc.).

Storage devices 112 of head unit 100 may include one or more computer-readable storage media. For example, storage devices 112 may be configured for long-term, as well as short-term storage of information, such as instructions, data, or other information used by head unit 100. In some examples, storage devices 112 may include non-volatile storage elements. Examples of such non-volatile storage elements include magnetic hard discs, optical discs, solid state discs, and/or the like. In other examples, in place of, or in addition to the non-volatile storage elements, storage devices 112 may include one or more so-called "temporary" memory devices, meaning that a primary purpose of these devices may not be long-term data storage. For example, the devices may comprise volatile memory devices, meaning that the devices may not maintain stored contents when the devices are not receiving power. Examples of volatile memory devices include random-access memories (RAM), dynamic random-access memories (DRAM), static random-access memories (SRAM), etc.

Computing system 120 of system 10 may be any suitable remote computing system, such as one or more desktop computers, laptop computers, mainframes, servers, cloud computing systems, virtual machines, etc., capable of sending and receiving information via network 111.

As shown in FIG .1, computing system 120 includes processors 122, which may be substantially similar to processors 102, COMM components 126, which may be substantially similar to COMM components 106, and storage devices 132, which may be substantially similar to storage devices 112. While described herein as being performed at least in part by computing system 120, any or all techniques of the present disclosure may be performed by one or more other devices, such as head unit 100. That is, in some examples, head unit 100 may be operable to perform one or more techniques of the present disclosure alone.

Head unit 100 may require permission from a user of a vehicle to perform an action, such as accessing the user's contact list (e.g., on a computing device connected to head unit 100), disclosing the user's information to make a third-party query, etc.). In general, head unit 100 may request the permission by, for example, displaying one or more GUI elements via display 104 of head unit 100. Responsive to the user providing user input (e.g., a tap, a long press, a swipe, etc.) via display 104, head unit 100 may perform the action corresponding to the user input (e.g., granting permission, denying permission, presenting additional information, etc.). However, this user experience may be dangerous because, to interact with display 104, the user may need to look at display 104 and remove at least one hand from the steering wheel of the vehicle, potentially distracting the user from a driving task.

In accordance with techniques of this disclosure, head unit 100 may include VUI module 116 configured to enable spoken human interaction with head unit 100 to respond to requests for permission (e.g., to access user personal data, to enable the usage of third-party services, etc.). For example, responsive to head unit 100 detecting that application 114 (e.g., a dialer application) has not been granted permission, VUI module 116 of head unit 100 may produce, via speakers 108 (e.g., speakers built into the vehicle, speakers integrated into a computing

device, portable speakers, etc.) an audio prompt requesting the required permission. In some examples, the audio prompt may identify the permission being requested. For example, if application 114 needs 'user device contacts setting' permission to access a user's contacts stored at a computing device (e.g., connected to head unit 100), the audio prompt emitted by VUI module 116 via speakers 108 may identify the permission being requested as the 'user device contacts setting' permission, thus giving a user more information to make a decision regarding the requested permission.

A user may respond to the audio prompt with an audio input in the form of human speech. Head unit 100 may receive the audio input from the user via microphones 110 (e.g., microphones built into the vehicle, microphones built into a computing device, portable microphones, etc.) and store the audio input data into an audio input repository 118. Head unit 100 may transmit, via COMM components 106, the audio input data stored in audio input repository 118 to computing system 120 via network 111. Computing system 120 may then store the audio input data in audio input repository 138.

A natural language understanding module 136 ("NLU module 136") of computing system 120 may be configured to analyze the audio input data stored in audio input repository 138 using speech recognition techniques to match the user's response with a valid input to the audio prompt provided by VUI module 116. A valid input may be any input with a semantic meaning that matches a semantic meaning of a predetermined response to the audio prompt, such as an input corresponding to a grant of permission, an input corresponding to a denial of permission, an input corresponding to a request for additional information, etc.

NLU module 136 may be a machine learning model trained to perform natural language processing tasks, such as converting speech to text, analyzing the text to determine its semantic

meaning, etc. In some examples, NLU module 136 may represent one more neural networks, such as one or more recurrent neural networks. In some instances, at least some of the nodes of a recurrent neural network may form a cycle. That is, NLU module 136 may pass or retain information from a previous portion of the input data sequence to a subsequent portion of the input data sequence through the use of recurrent or directed cyclical node connections, where the input data sequence includes words in a sentence for natural language processing, speech detection or processing, etc.

Responsive to NLU module 136 determining the input to which the audio input corresponds, computing system 120 may transmit an indication of the corresponding input to head unit 100 via network 111. Head unit 100 may then perform the action associated with the corresponding input. For example, if NLU module 136 determines that the audio input corresponds to granting application 114 permission to access a user's contacts stored at a computing device (e.g., connected to head unit 100), head unit 100 may grant application 114 permission to do so. In another example, if NLU module 136 determines that the audio input corresponds to denying application 114 permission to access a user's contacts stored at a computing device (e.g., connected to head unit 100), head unit 100 may deny application 114 permission to do so. In yet another example, if NLU module 136 cannot determine an input to which audio input corresponds, VUI module 116 may repeat the audio prompt to the user to receive and process a new audio input.
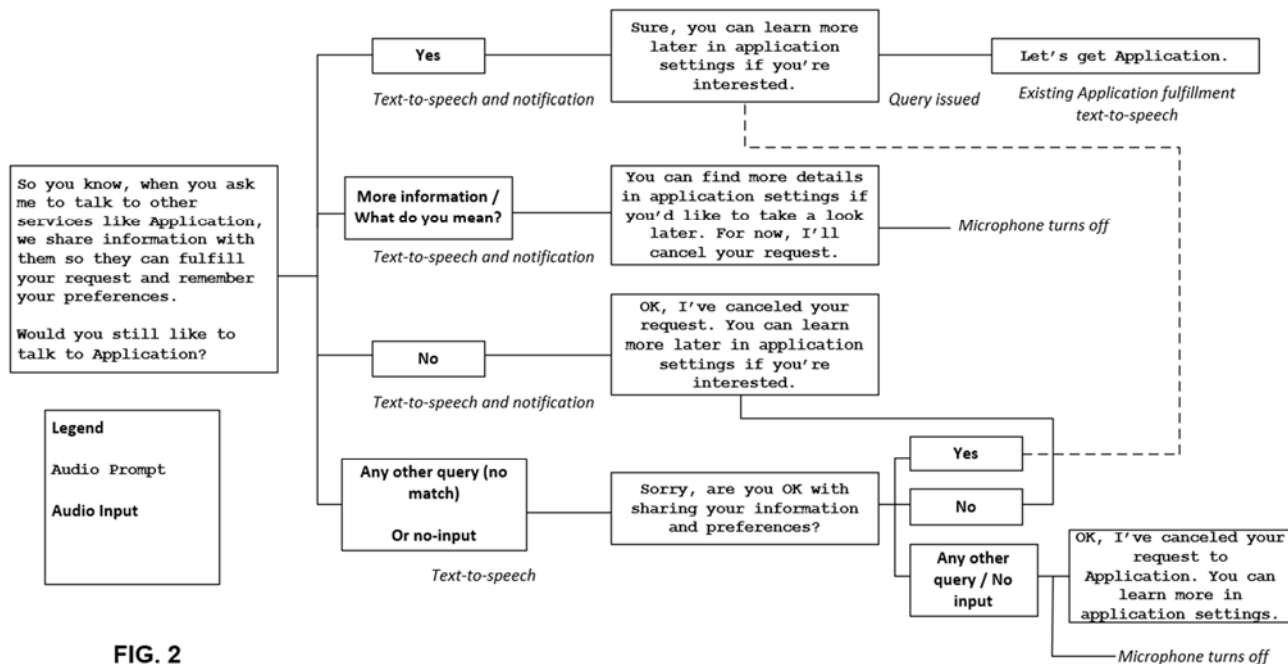
FIG. 2

FIG. 2 is a flow diagram illustrating an example dialogue between VUI module 116 and a user, where the blue text represents audio prompts from VUI module 116, and the green text represents audio input from the user. As shown in FIG. 2, VUI module 116 may explain, by emitting sound via speakers 108, the details of the permission being requested by, for example, application 114, which in the example of FIG. 2 is a third-party application, and ask the user to grant permission. The style of the audio prompt provided by VUI module 116 may be personalized and colloquial to encourage a user-friendly dialogue. VUI module 116 may receive audio input (e.g., "yes," "more information," "no," any other query, etc.) from the user via microphones 110. Head unit 100 may then transmit the audio input data to computing system 120 via network 111 for NLU module 136 to match the user's response with a corresponding (valid) input to the audio prompt from the user. Upon completion thereof, computing system 120

may transmit an indication of the corresponding input to head unit 100 via network 111. Head unit 100 may then perform an action associated with the corresponding input.

For example, if a user says, "sure," to the audio prompt from VUI module 116 requesting permission on behalf of application 114, NLU module 136 may process the audio input data associated with that audio input and determine that the semantic meaning of "sure" matches the semantic meaning of the predetermined response "yes." NLU module 136 may transmit an indication that the corresponding input is "yes," and VUI module 116 may convert a predetermined text response associated with the corresponding input "yes" into speech (e.g., outputted by speakers 108). In the example of FIG. 2, the predetermined text response is "sure, you can learn more later in assistant settings if you're interested." As shown in the example of FIG. 2, VUI module 116 may also provide a notification (e.g., via display 104) associated with the corresponding input "yes." VUI module 116 may perform the action associated with the corresponding input "yes" by granting application 114 permission to access user information to fulfill requests and save preferences.

In another example, if a user says, "I don't know. Why does the application want my information?" NLU module 136 may determine that the semantic meaning of that audio data (e.g., after being converted to text) matches the semantic meaning of the predetermined response "More information/what do you mean?" NLU module 136 may transmit an indication of that corresponding input, and VUI module 116 may output a text-to-speech response via speakers 108, such as "You can find more details in application settings if you'd like to take a look later. For now, I'll cancel your request." Head unit 100 may then perform the action associated with "More information/what do you mean?" by cancelling the request. In yet another example, if the user says, "nope," NLU module 136 may determine that the semantic meaning of "nope"

matches the semantic meaning of the predetermined response "no," resulting in VUI module 116 outputting the text-to-speech response, "OK, I've canceled your request. You can learn more later in application settings if you're interested." Head unit 100 may then cancel the request.

In yet another example, if a user provides an audio input with no corresponding input (e.g., the user says something unrelated to the audio prompt) or no input (e.g., the user is silent), NLU module 136 may determine that the semantic meaning of the audio input matches the semantic meaning of "Any other query (no match) / no-input." In this case, VUI module 116 may output another audio prompt, such as "Sorry, are you OK with sharing your information and preferences?" The user may respond to the audio prompt with an audio input similar to the ones described in previous examples (e.g., "yes," "no," any other query, no input, etc.). If the user responds "yes," then head unit 100 fulfills the request for permission and outputs dialogue like "Let's get the application." If the user responds "no," then head unit 100 denies the request and may state that the request has been canceled. If the audio data corresponds to any other query or no input, then head unit 100 may cancel the request, and VUI module 116 may output "OK, I've canceled your request to Application. You can learn more in application settings."

Responsive to VUI module 116 outputting terminating dialogue (e.g., "Let's get the application," "OK, I've canceled your request to Application. You can learn more in application settings," etc.) head unit 100 may stop receiving or processing audio data from microphone 110. For example, head unit 100 may cause microphone 110 to not actively detect audio by disabling microphone 110, powering off microphone 110, etc. In some examples, head unit 100 may cause microphone 110 to stop providing the stream of audio data to processors 102.

One or more advantages of the techniques described in this disclosure include providing a user a hands-free, eyes-free user experience for controlling the granting of permissions via a

VUI, which may be particularly beneficial in vehicle settings in which the user is operating the vehicle, as such a user experience may reduce distractions and thereby promote safety. Another advantage includes providing a user the opportunity to address the granting or denying of permissions in real-time or just-in-time (e.g., as opposed to notifying the user that an application was granted permission to perform an action without the user's consent), thus promoting the user's interests in data privacy, network usage, etc.

It is noted that the techniques of this disclosure may be combined with any other suitable technique or combination of techniques. As one example, the techniques of this disclosure may be combined with the techniques described in U.S. Patent Application Publication No. 2020/0099538A1. In another example, the techniques of this disclosure may be combined with the techniques described in U.S. Patent Application Publication No. 2020/0143812A1. In yet another example, the techniques of this disclosure may be combined with the techniques described in U.S. Patent Application Publication No. 2020/0221420A1. In yet another example, the techniques of this disclosure may be combined with the techniques described in U.S. Patent Application Publication No. 2016/0164881A1.