July 2021

PROVIDE VEHICLE MICROPHONE FOR A THIRD-PARTY APPLICATION

Rafael Lima

Ozge Kirimlioglu

Christian Gonzalez

PROVIDE VEHICLE MICROPHONE FOR A THIRD-PARTY APPLICATION

ABSTRACT

When a user is driving 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.), a computing device (e.g., a smartphone, a tablet computer, smartglasses, a smartwatch, a portable gaming system, a laptop computer, etc.) may be located in the user's pocket, purse, etc., potentially worsening the quality of the audio data recorded by a microphone built into the computing device. To address this issue, the computing device may enable applications installed at the computing device to use microphones built into the vehicle (which potentially records higher quality audio data than microphones built into the computing device). An operating system executing at the computing device may provide a library (referred to herein as a "voice template") accessible by the applications installed at the computing device. The applications may request access to the vehicle microphones using the voice template and may receive audio data from the vehicle microphones using the voice template. As a result, applications executing at the computing device may receive higher quality audio data, which may improve the accuracy and reliability of voice input processing.

DESCRIPTION

FIG. 1 below is a conceptual diagram of a system 10 including a head unit 100 of a vehicle (e.g., an automobile, a 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 device 120 (e.g., a smartphone, a tablet computer, smartglasses, a smartwatch, a portable gaming system, a laptop computer, etc.). 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 microphones 108, one or more speakers 110, and one or more storage devices 112. As further shown in FIG. 2, computing device 120 includes one or more processors 122, a display 124, one or more communication components 126 ("COMM components 126"), one or more microphones 128, one or more speakers 130, and one or more storage devices 132.

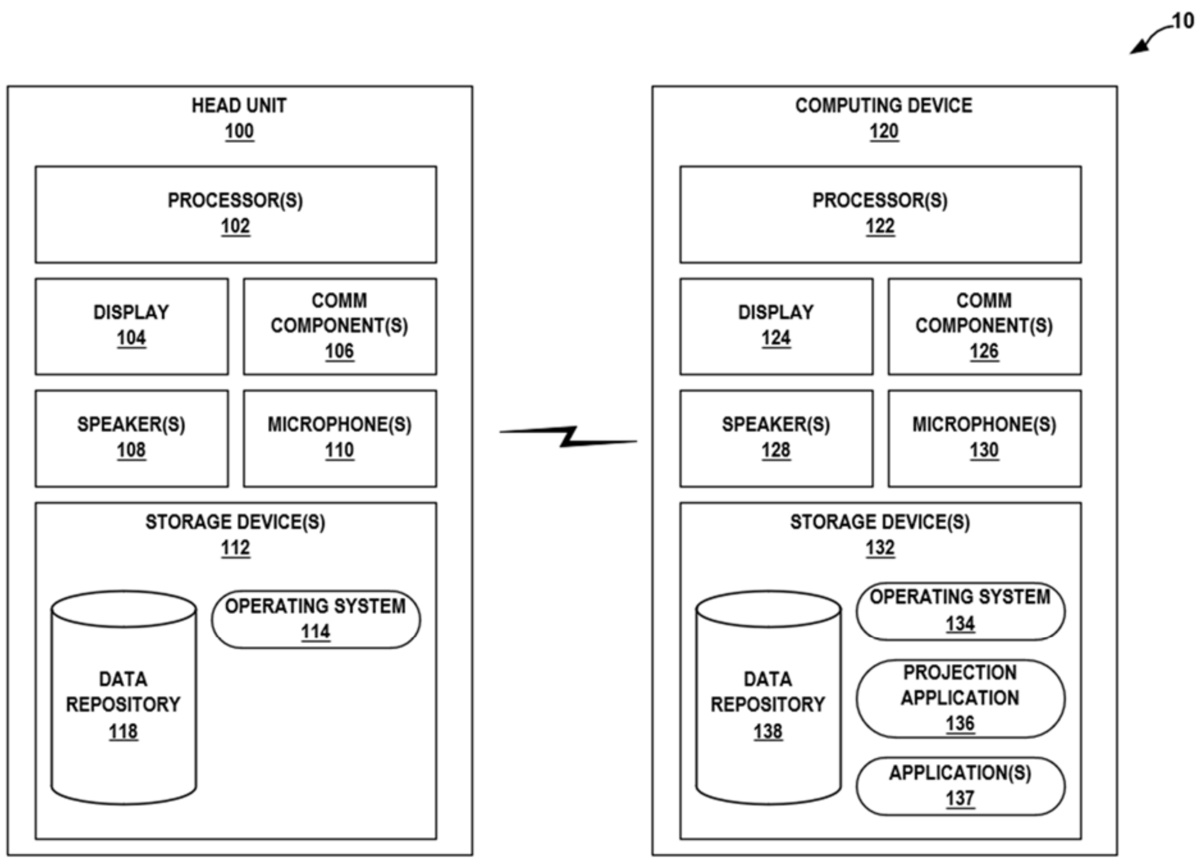

FIG. 1

One or more components (e.g., processors 102, display 104, COMM components 106, microphones 108, speakers 110, storage devices 112, etc.) of head unit 100 may be substantially similar to one or more components (e.g., processors 122, display 124, COMM components 126, microphones 128, speakers 130, storage devices 132, etc.) of second computing device 120. As such, the description of one may apply equally to the other except for any differences described here.

Head unit 100 of system 10 may operate to assist, inform, entertain, or otherwise provide for interactions with one or more occupants of a vehicle. Head unit 100 may represent an integrated head unit that provides a user interface (UI), such as 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 configured to establish a session with a computing device to permit data exchange. In some examples, an occupant of the vehicle in which vehicle head unit 100 is located may connect computing device 120 to head unit 100 to project (or otherwise, cast or stream) a UI to head unit 100. For instance, a UI model in head unit 100 may be a thin client that supports projection of a GUI from computing device 120.

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. Processors 102 may retrieve and execute instructions stored by storage devices 112 that cause processors 102 to perform the operations described in this disclosure.

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, 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, 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 (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, etc. Examples of volatile memory devices include random-access memories (RAM), dynamic random-access memories (DRAM), static random-access memories (SRAM), etc.

As shown in FIG. 1, storage devices 112 may include an operating system 114 ("OS 114") that provides an execution environment for one or more applications, such as applications

137. OS 114 may represent a multi-threaded operating system or a single-threaded operating system with which a projection application 136 may interface to access hardware of head unit 100. OS 114 may include a kernel that facilitates access to the underlying hardware of head unit 100, where kernel may present a number of different interfaces (e.g., application programmer interfaces – APIs) that projection application 136 may invoke to access the underlying hardware of head unit 100.

Projection application 136 may represent an application that provides a bridge between applications 137 and facilitates projection of UIs, such as GUIs, VUIs, etc., for applications 137 to head unit 100. In some examples, projection application 136 may provide, manage, update, and/or control the UIs. For instance, projection application 136 may provide via display 104 a GUI that, for example, includes information to the user in the form of text, images, etc. Projection application 136 may generate graphical elements to, for example, satisfy driver distraction standards and accommodate a variety of car screen factors and input modalities. Actuation of graphical elements may invoke corresponding functions of applications 137.

In general, application 137 (e.g., a navigation application, a music streaming application, etc.) executing at computing device 120 may process a voice input from a user to perform a corresponding function. Application 137 may receive audio data associated with the voice input via microphone 128 (which is built into computing device 120). However, when the user is driving a vehicle (e.g., one into which head unit 100 is built), computing device 120 may be located in the user's pocket, purse, etc., potentially worsening the quality of the audio data recorded by microphone 128. Thus, it may be advantageous for application 137 to be able to receive audio data associated with the voice input via a microphone other than microphone 128.

In accordance with techniques of this disclosure, application 137 may request access to microphones 108 using a voice template (e.g., a library provided by an operating system of computing device 120 accessible by application 137) and may receive audio data from microphones 108 using the voice template. As a result, application 137 may receive higher quality audio data, which may improve the accuracy and reliability of voice input processing. In some examples, developers of application 137 (which may be third-party applications) may define the way the voice template manages microphones 108. For instance, application 137 may be configured to duck, stop, pause, etc., recording of audio data to improve the user experience of application 137.

As an example, application 137 may be in a speaking state during which application 137 emits, for example, an audio prompt to the user via speakers 110, speakers 130, etc. Responsive to termination of the speaking state (due to, e.g., completion of the audio prompt, occurrence of a pre-determined length of silence, etc.), application 137 may transition into a listening state during which application 137 uses a voice template to activate microphones 108.

The voice template may include a variety of features to limit access to microphones 108. As an example, the voice template may disable recording via microphones 108 during a telephone call (e.g., occurring at computing device 120, which may be a smartphone). In another example, the voice template may enable only one application to access microphones 108 at a time. In yet another example, the voice template may require that application 137 be in the foreground for application 137 to access microphones 108. In general, application 137 may be required to have permission (e.g., granted by the user of computing device 120) to record audio (e.g., using microphones 108, microphones 128, etc.). For example, to ensure that application

137 is authorized to obtain the audio data, the voice template may check whether application 137 possesses permission from the user of computing device to access microphones 108.

Responsive to verifying that application 137 possesses the required permission, the voice template may, via projection application 136, interface with OS 114 to cause microphones 108 to record audio data (e.g., voice commands) from the user. OS 114 may then store the audio data in data repository 118. While application 137 is in a listening state, projection application 136 may change the visual appearance of a GUI of the voice template to indicate to the user that microphones 108 are active and thus recording audio data.

During the listening session or in response to initiating the listening session, OS 114 may send a callback containing the audio data (stored in data repository 118) to computing device 120 via a binder (e.g., a cross-process communication mechanism). As an example, the callback may use a file descriptor representing an open file, an open socket, or another source or sink of bytes to create an input stream for the bytes. In some instances, head unit 100 may stream the bytes to computing device 120 (that computing device 120 stores in a data repository 138) for application 137 to process in substantially real-time.

In some examples, application 137 may retrieve a uniform resource identifier (URI) (e.g., a unique sequence of characters that identifies a logical or physical resource) corresponding to the audio data from head unit 100. For example, head unit 100 may send the URI in a callback to application 137, and application 137 may in turn provide the URI to head unit 100 to request access to the corresponding audio data. Responsive to OS 114 locating the corresponding audio data in data repository 118, projection application 136 may initiate creation of the input stream to send the corresponding audio data to computing device 120.

Application 137 may terminate the listening state (due to, e.g., completion of the audio command, occurrence of a pre-determined length of silence, etc.). Because application 137 is no longer in the listening state, head unit 100 may deactivate microphones 108 such that microphones 108 are no longer recording audio data. Additionally, head unit 100 may close the input stream to computing device 120 for application 137 to stop the sending of audio data to computing device 120. Application 137 may then transition to a thinking state during which application 137 converts, via operating system 134, the bytes into audio data.

One or more advantages of the techniques described in this disclosure include enabling applications, such as third-party applications, executing at a computing device to access microphones built into a vehicle, which may record higher quality audio data than microphones built into the computing device. In this way, the techniques may improve the accuracy and reliability of voice input processing. Further, the voice template may include a variety of features to protect user privacy, such as disabling recording via microphones built into the vehicle during a telephone call to limit access to the microphones.

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. 2017/0228311A1. In another example, the techniques of this disclosure may be combined with the techniques described in “Automotive Audio,” Google Android. In yet another example, the techniques of this disclosure may be combined with the techniques described in “Getting Microphone Audio,” Smart Device Link. In yet another example, the techniques of this disclosure may be combined with the techniques described in “AudioRecord,” Google Android. In yet another example, the techniques of this disclosure may be combined with the techniques described in “CarInputService.java,” Google Git. In yet another example, the techniques of this disclosure may be combined with the techniques described in Eriksson et al., “On the road with third-party apps,” Chalmers University of Technology, 2018.