

SUB-VOCAL SPEECH DETECTION ON SMART GLASSES

BACKGROUND

[0001] Users are increasingly wearing smart glasses. A user may give voice commands to a processor of the smart glasses, e.g., to perform certain functions. However, detecting and
5 identifying the voice commands may be difficult.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] Illustrative embodiments are described in detail below with reference to the following figures.

[0003] FIG. 1 illustrates an example implementation of a pair of smart glasses in accordance
10 with some embodiments.

[0004] FIG. 2 is a flowchart of an example process for determining sub-vocal or quiet speech in accordance with some embodiments.

[0005] FIG. 3 is a simplified block diagram of an example of a computing system that may be implemented as part of a user device according to certain embodiments.

15 [0006] The figures depict embodiments of the present disclosure for purposes of illustration only. One skilled in the art will readily recognize from the following description that alternative embodiments of the structures and methods illustrated may be employed without departing from the principles, or benefits touted, of this disclosure.

20 [0007] In the appended figures, similar components and/or features may have the same reference label. Further, various components of the same type may be distinguished by following the reference label by a dash and a second label that distinguishes among the similar components. If only the first reference label is used in the specification, the description is applicable to any one of the similar components having the same first reference label irrespective of the second reference label.

DETAILED DESCRIPTION

[0008] Users are increasingly wearing smart glasses. A user may give voice commands to a processor of the smart glasses, e.g., to perform certain functions. For example, voice commands may be used to activate an artificial intelligence (AI) assistant, use speech-to-text for, e.g., sending
5 messages, activate particular functions (e.g., to initiate playback of certain media content, to take a picture, etc.) using the smart glasses, or the like. However, a user may be hesitant to give voice commands while in public, for example, due to embarrassment or a desire for the content of the voice command to remain private (e.g., when using speech-to-text to send a message).

[0009] Disclosed herein are techniques for detecting and/or identifying sub-vocal or quiet speech
10 using one or more sensors disposed in or embedded in a portion of a frame of a pair of smart glasses. The one or more sensors may include electromyography (EMG) electrodes configured to sense muscle contractions, inertial measurement unit (IMU) sensors, microphones, etc. In some implementations, a microphone may be a contact microphone which senses audio vibrations through contact with solid objects, which may include portions of the wearer's face and/or a
15 portion of the frame that is in contact with the wearer's face. As used herein, "quiet speech" may include whispered speech where there is no vibration of the talker's vocal cords, speech below a particular volume (e.g., below a typical conversation level of 60 decibels, below 40 decibels, below 30 decibels, etc.). In some cases, "quiet speech" may refer to speech having a signal-to-noise ratio (SNR) below a predetermined threshold with respect to background noise (e.g., below 10 dB SNR,
20 below 5 dB SNR, below 0 dB SNR, etc.). As used herein, "sub-vocal" speech may involve speech that does not involve audible sound but involves muscle contractions (e.g., of the muscles surrounding the jaw, the ear, etc.).

[0010] FIG. 1 illustrates an example pair of smart glasses in accordance with some embodiments. As illustrated, smart glasses 100 includes sensors, including EMG sensor 102, and
25 sensors 104, 106, and 108. Note that the sensors are disposed in the frame of smart glasses 100. EMG sensors (such as EMG sensor 102) may receive and/or detect electrical signals associated with contractions of facial muscles of the wearer, which may be associated with production of speech (whether audible, quiet, or sub-vocal). Note that the sensors may additionally or

alternatively include one or more microphones, one or more IMU sensors, etc., in addition to one or more EMG sensors.

[0011] FIG. 2 is a flowchart of an example process 200 for detecting or identifying sub-vocal or quiet speech in accordance with some embodiments. Blocks of process 200 may be executed by one or more processors of a pair of smart glasses. An example of a computing device that may be used to implement a pair of smart glasses is shown in and described below in connection with FIG. 3. In some implementations, blocks of process 300 may be executed in an order other than what is shown in FIG. 2. In some embodiments, two or more blocks of process 300 may be executed substantially in parallel. In some implementations, one or more blocks of process 300 may be omitted.

[0012] Process 200 can begin at 202 by receiving signals from one or more sensors disposed in a portion of a frame of a pair of smart glasses. The sensors may include one or more microphones, one or more IMU sensors, one or more EMG sensors, etc. The sensors may be disposed in portions of the frame that are configured to be in contact with a wearer's head or face, such as on a hook that rests behind the wearer's ears, a nose bridge, a side frame configured to rest on a side of the wearer's face, etc.

[0013] At 204, process 200 can determine sub-vocal or quiet speech spoken by the wearer of the pair of smart glasses based on the received signals. For example, process 200 may identify one or more speech tokens from a closed set of speech tokens. The speech tokens may include commonly spoken words or phrases (e.g., "yes," "no," "play," "stop," "send text," etc.). As another example, in some implementations, process 200 may identify speech from an open set. In some implementations, process 200 may determine the sub-vocal or quiet speech by providing the received signals, or representations of the received signals, to a trained machine learning model. In instances in which the trained machine learning model is configured to identify speech tokens from a closed set, the machine learning model may be a classification type algorithm. In some embodiments, any sort of pre-processing or feature extraction may be performed on the signals prior to providing extracted features to the trained machine learning model.

[0014] Note that, in some embodiments, signals from different types of sensors may be combined. For example, signals from an EMG sensor and a microphone may be combined. In

some implementations, signals may be obtained from multiple types of sensors continuously, and signals from a particular type of sensor may be selected based on an identified volume of speech. For example, signals from a microphone may be selected responsive to determining the speech is whispered, or otherwise quiet but audible. As another example, signals from an EMG sensor may be selected responsive to determining the speech is sub-vocal. Note that in some embodiments, EMG signals may be used to decode speech of any volume (or sub-vocal speech). For example, EMG signals may be used to disambiguate whispered speech, or speech having a relatively low SNR.

[0015] In some embodiments, sensor signals may be used to perform other functions, such as eye tracking. For example, EMG signals may be used to obtain an electro-oculography (EOG) signal indicative of where the user is looking.

[0016] FIG. 3 is a simplified block diagram of an example of a computing system 300 for implementing some of the examples described herein. For example, in some embodiments, computing system may be used to implement a pair of smart glasses that implements the blocks of processes 200 of FIG. 2. In the illustrated example, computing system 300 may include one or more processor(s) 310 and a memory 320. Processor(s) 310 may be configured to execute instructions for performing operations at a number of components, and can be, for example, a general-purpose processor or microprocessor suitable for implementation within a portable electronic device. Processor(s) 310 may be communicatively coupled with a plurality of components within computing system 300. To realize this communicative coupling, processor(s) 310 may communicate with the other illustrated components across a bus 340. Bus 340 may be any subsystem adapted to transfer data within computing system 300. Bus 340 may include a plurality of computer buses and additional circuitry to transfer data.

[0017] Memory 320 may be coupled to processor(s) 310. In some embodiments, memory 320 may offer both short-term and long-term storage and may be divided into several units. Memory 320 may be volatile, such as static random access memory (SRAM) and/or dynamic random access memory (DRAM) and/or non-volatile, such as read-only memory (ROM), flash memory, and the like. Furthermore, memory 320 may include removable storage devices, such as secure digital (SD) cards. Memory 320 may provide storage of computer-readable instructions, data structures,

program modules, and other data for computing system 300. In some embodiments, memory 320 may be distributed into different hardware modules. A set of instructions and/or code might be stored on memory 320. The instructions might take the form of executable code that may be executable by computing system 300, and/or might take the form of source and/or installable code, which, upon compilation and/or installation on computing system 300 (e.g., using any of a variety of generally available compilers, installation programs, compression/decompression utilities, etc.), may take the form of executable code.

[0018] In some embodiments, memory 320 may store a plurality of application modules 322 through 324, which may include any number of applications. Examples of applications may include gaming applications, conferencing applications, video playback applications, or other suitable applications. The applications may include a depth sensing function or eye tracking function. Application modules 322-324 may include particular instructions to be executed by processor(s) 310. In some embodiments, certain applications or parts of application modules 322-324 may be executable by other hardware modules 380. In certain embodiments, memory 320 may additionally include secure memory, which may include additional security controls to prevent copying or other unauthorized access to secure information.

[0019] In some embodiments, memory 320 may include an operating system 325 loaded therein. Operating system 325 may be operable to initiate the execution of the instructions provided by application modules 322-324 and/or manage other hardware modules 380 as well as interfaces with a wireless communication subsystem 330 which may include one or more wireless transceivers. Operating system 325 may be adapted to perform other operations across the components of computing system 300 including threading, resource management, data storage control and other similar functionality.

[0020] Wireless communication subsystem 330 may include, for example, an infrared communication device, a wireless communication device and/or chipset (such as a Bluetooth® device, an IEEE 802.11 device, a Wi-Fi device, a WiMax device, cellular communication facilities, etc.), and/or similar communication interfaces. Computing system 300 may include one or more antennas 334 for wireless communication as part of wireless communication subsystem 330 or as a separate component coupled to any portion of the system. Depending on desired functionality,

wireless communication subsystem 330 may include separate transceivers to communicate with base transceiver stations and other wireless devices and access points, which may include communicating with different data networks and/or network types, such as wireless wide-area networks (WWANs), wireless local area networks (WLANs), or wireless personal area networks (WPANs). A WWAN may be, for example, a WiMax (IEEE 802.17) network. A WLAN may be, for example, an IEEE 802.11x network. A WPAN may be, for example, a Bluetooth network, an IEEE 802.7x, or some other types of network. The techniques described herein may also be used for any combination of WWAN, WLAN, and/or WPAN. Wireless communications subsystem 330 may permit data to be exchanged with a network, other computer systems, and/or any other devices described herein. Wireless communication subsystem 330 may include a means for transmitting or receiving data, such as identifiers of HMD devices, position data, a geographic map, a heat map, photos, or videos, using antenna(s) 334 and wireless link(s) 332. Wireless communication subsystem 330, processor(s) 310, and memory 320 may together comprise at least a part of one or more of a means for performing some functions disclosed herein.

15 **[0021]** Embodiments of computing system 300 may also include one or more sensors 390. Sensor(s) 390 may include, for example, an EMG sensor, an image sensor, an accelerometer, a force sensor, a hydrostatic pressure sensor, a temperature sensor, a proximity sensor, a magnetometer, a gyroscope, an inertial sensor (e.g., a module that combines an accelerometer and a gyroscope), an ambient light sensor, or any other similar module operable to provide sensory output and/or receive sensory input, such as a depth sensor or a position sensor. For example, in some implementations, sensor(s) 390 may include one or more inertial measurement units (IMUs) and/or one or more position sensors. An IMU may generate calibration data indicating an estimated position of a device, based on measurement signals received from one or more of the position sensors. A position sensor may generate one or more measurement signals in response to motion of a device. Examples of the position sensors may include, but are not limited to, one or more accelerometers, one or more gyroscopes, one or more magnetometers, another suitable type of sensor that detects motion, a type of sensor used for error correction of the IMU, or some combination thereof. The position sensors may be located external to the IMU, internal to the IMU, or some combination thereof. At least some sensors may use a structured light pattern for sensing.

20

25

[0022] Computing system 300 may include a display module 330. Display module 330 may be a near-eye display, and may graphically present information, such as images, videos, and various instructions, from computing system 300 to a user. Such information may be derived from one or more application modules 322-324, virtual reality engine 326, one or more other hardware modules
5 380, a combination thereof, or any other suitable means for resolving graphical content for the user (e.g., by operating system 325). Display module 360 may use liquid crystal display (LCD) technology, light-emitting diode (LED) technology (including, for example, OLED, ILED, μ LED, AMOLED, TOLED, etc.), light emitting polymer display (LPD) technology, or some other display technology.

10 **[0023]** Computing system 300 may include a user input/output module 370. User input/output module 370 may allow a user to send action requests to computing system 300. An action request may be a request to perform a particular action. For example, an action request may be to start or end an application or to perform a particular action within the application. User input/output module 370 may include one or more input devices. Example input devices may include a
15 touchscreen, a touch pad, microphone(s), button(s), dial(s), switch(es), a keyboard, a mouse, a game controller, or any other suitable device for receiving action requests and communicating the received action requests to computing system 300. In some embodiments, user input/output module 370 may provide haptic feedback to the user in accordance with instructions received from computing system 300. For example, the haptic feedback may be provided when an action request
20 is received or has been performed.

[0024] Computing system 300 may include a camera 350 that may be used to take photos or videos. Camera 350 may be configured to take photos or videos of the user. Camera 350 may also be used to take photos or videos of the environment, for example, for VR, AR, or MR applications. Camera 350 may include, for example, a complementary metal-oxide-
25 semiconductor (CMOS) image sensor with a few millions or tens of millions of pixels. In some implementations, camera 350 may include two or more cameras that may be used to capture 3-D images.

[0025] In some embodiments, computing system 300 may include a plurality of other hardware modules 380. Each of other hardware modules 380 may be a physical module within computing

system 300. While each of other hardware modules 380 may be permanently configured as a structure, some of other hardware modules 380 may be temporarily configured to perform specific functions or temporarily activated. Examples of other hardware modules 380 may include, for example, an audio output and/or input module (*e.g.*, a microphone or speaker), a near field communication (NFC) module, a rechargeable battery, a battery management system, a wired/wireless battery charging system, etc. In some embodiments, one or more functions of other hardware modules 380 may be implemented in software.

[0026] In some embodiments, memory 320 of computing system 300 may also store a virtual reality engine 326. Virtual reality engine 326 may execute applications within computing system 300 and receive position information, acceleration information, velocity information, predicted future positions, or some combination thereof from the various sensors. In some embodiments, the information received by virtual reality engine 326 may be used for producing a signal (*e.g.*, display instructions) to display module 360. For example, if the received information indicates that the user has looked to the left, virtual reality engine 326 may generate content that mirrors the user's movement in a virtual environment. Additionally, virtual reality engine 326 may perform an action within an application in response to an action request received from user input/output module 370 and provide feedback to the user. The provided feedback may be visual, audible, or haptic feedback. In some implementations, processor(s) 310 may include one or more GPUs that may execute virtual reality engine 326.

[0027] In various implementations, the above-described hardware and modules may be implemented on a single device or on multiple devices that can communicate with one another using wired or wireless connections. For example, in some implementations, some components or modules, such as GPUs, virtual reality engine 326, and applications (*e.g.*, tracking application), may be implemented on two or more paired or connected devices.

[0028] In alternative configurations, different and/or additional components may be included in computing system 300. Similarly, functionality of one or more of the components can be distributed among the components in a manner different from the manner described above. For example, in some embodiments, computing system 300 may be modified to include other system environments, such as an AR system environment and/or an MR environment.

[0029] Embodiments disclosed herein may be used to implement components of an artificial reality system or may be implemented in conjunction with an artificial reality system. Artificial reality is a form of reality that has been adjusted in some manner before presentation to a user, which may include, for example, a virtual reality, an augmented reality, a mixed reality, a hybrid reality, or some combination and/or derivatives thereof. Artificial reality content may include completely generated content or generated content combined with captured (*e.g.*, real-world) content. The artificial reality content may include video, audio, haptic feedback, or some combination thereof, and any of which may be presented in a single channel or in multiple channels (such as stereo video that produces a three-dimensional effect to the viewer). Additionally, in some embodiments, artificial reality may also be associated with applications, products, accessories, services, or some combination thereof, that are used to, for example, create content in an artificial reality and/or are otherwise used in (*e.g.*, perform activities in) an artificial reality. The artificial reality system that provides the artificial reality content may be implemented on various platforms, including an HMD connected to a host computer system, a standalone HMD, a mobile device or computing system, or any other hardware platform capable of providing artificial reality content to one or more viewers.

[0030] The methods, systems, and devices discussed above are examples. Various embodiments may omit, substitute, or add various procedures or components as appropriate. For instance, in alternative configurations, the methods described may be performed in an order different from that described, and/or various stages may be added, omitted, and/or combined. Also, features described with respect to certain embodiments may be combined in various other embodiments. Different aspects and elements of the embodiments may be combined in a similar manner. Also, technology evolves and, thus, many of the elements are examples that do not limit the scope of the disclosure to those specific examples.

[0031] Specific details are given in the description to provide a thorough understanding of the embodiments. However, embodiments may be practiced without these specific details. For example, well-known circuits, processes, systems, structures, and techniques have been shown without unnecessary detail in order to avoid obscuring the embodiments. This description provides example embodiments only, and is not intended to limit the scope, applicability, or configuration of the invention. Rather, the preceding description of the embodiments will provide those skilled

in the art with an enabling description for implementing various embodiments. Various changes may be made in the function and arrangement of elements without departing from the spirit and scope of the present disclosure.

[0032] Also, some embodiments were described as processes depicted as flow diagrams or block diagrams. Although each may describe the operations as a sequential process, many of the operations may be performed in parallel or concurrently. In addition, the order of the operations may be rearranged. A process may have additional steps not included in the figure. Furthermore, embodiments of the methods may be implemented by hardware, software, firmware, middleware, microcode, hardware description languages, or any combination thereof. When implemented in software, firmware, middleware, or microcode, the program code or code segments to perform the associated tasks may be stored in a computer-readable medium such as a storage medium. Processors may perform the associated tasks.

[0033] It will be apparent to those skilled in the art that substantial variations may be made in accordance with specific requirements. For example, customized or special-purpose hardware might also be used, and/or particular elements might be implemented in hardware, software (including portable software, such as applets, etc.), or both. Further, connection to other computing devices such as network input/output devices may be employed.

[0034] With reference to the appended figures, components that can include memory can include non-transitory machine-readable media. The term “machine-readable medium” and “computer-readable medium” may refer to any storage medium that participates in providing data that causes a machine to operate in a specific fashion. In embodiments provided hereinabove, various machine-readable media might be involved in providing instructions/code to processing units and/or other device(s) for execution. Additionally or alternatively, the machine-readable media might be used to store and/or carry such instructions/code. In many implementations, a computer-readable medium is a physical and/or tangible storage medium. Such a medium may take many forms, including, but not limited to, non-volatile media, volatile media, and transmission media. Common forms of computer-readable media include, for example, magnetic and/or optical media such as compact disk (CD) or digital versatile disk (DVD), punch cards, paper tape, any other physical medium with patterns of holes, a RAM, a programmable read-only memory (PROM), an

erasable programmable read-only memory (EPROM), a FLASH-EPROM, any other memory chip or cartridge, a carrier wave as described hereinafter, or any other medium from which a computer can read instructions and/or code. A computer program product may include code and/or machine-executable instructions that may represent a procedure, a function, a subprogram, a program, a routine, an application (App), a subroutine, a module, a software package, a class, or any combination of instructions, data structures, or program statements.

[0035] Those of skill in the art will appreciate that information and signals used to communicate the messages described herein may be represented using any of a variety of different technologies and techniques. For example, data, instructions, commands, information, signals, bits, symbols, and chips that may be referenced throughout the above description may be represented by voltages, currents, electromagnetic waves, magnetic fields or particles, optical fields or particles, or any combination thereof.

[0036] Terms, “and” and “or” as used herein, may include a variety of meanings that are also expected to depend at least in part upon the context in which such terms are used. Typically, “or” if used to associate a list, such as A, B, or C, is intended to mean A, B, and C, here used in the inclusive sense, as well as A, B, or C, here used in the exclusive sense. In addition, the term “one or more” as used herein may be used to describe any feature, structure, or characteristic in the singular or may be used to describe some combination of features, structures, or characteristics. However, it should be noted that this is merely an illustrative example and claimed subject matter is not limited to this example. Furthermore, the term “at least one of” if used to associate a list, such as A, B, or C, can be interpreted to mean any combination of A, B, and/or C, such as A, AB, AC, BC, AA, ABC, AAB, AABBBCCC, etc.

[0037] Further, while certain embodiments have been described using a particular combination of hardware and software, it should be recognized that other combinations of hardware and software are also possible. Certain embodiments may be implemented only in hardware, or only in software, or using combinations thereof. In one example, software may be implemented with a computer program product containing computer program code or instructions executable by one or more processors for performing any or all of the steps, operations, or processes described in this disclosure, where the computer program may be stored on a non-transitory computer readable

medium. The various processes described herein can be implemented on the same processor or different processors in any combination.

5 [0038] Where devices, systems, components or modules are described as being configured to perform certain operations or functions, such configuration can be accomplished, for example, by designing electronic circuits to perform the operation, by programming programmable electronic circuits (such as microprocessors) to perform the operation such as by executing computer instructions or code, or processors or cores programmed to execute code or instructions stored on a non-transitory memory medium, or any combination thereof. Processes can communicate using a variety of techniques, including, but not limited to, conventional techniques for inter-process
10 communications, and different pairs of processes may use different techniques, or the same pair of processes may use different techniques at different times.

[0039] The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. It will, however, be evident that additions, subtractions, deletions, and other modifications and changes may be made thereunto without departing from the broader spirit
15 and scope as set forth in the claims. Thus, although specific embodiments have been described, these are not intended to be limiting. Various modifications and equivalents are within the scope of the following claims.