

(43) International Publication Date
4 September 2014 (04.09.2014)(51) International Patent Classification:
G06F 17/00 (2006.01)(21) International Application Number:
PCT/US2014/013070(22) International Filing Date:
24 January 2014 (24.01.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/777,255 26 February 2013 (26.02.2013) US(71) Applicants: **HONDA MOTOR CO., LTD.** [JP/JP]; 2-1-1 Minami Aoyama, Minato-ku, Tokyo, 107-8556 (JP). **EDGE3 TECHNOLOGIES, INC.** [US/US]; 688 W 1st St., Tempe, AZ 85281 (US).(72) Inventors: **VAGHEFINAZARI, Pedram**; C/o Honda Patents & Technologies North, America, Llc, 21001 State Route 739, Raymond, OH 43067-9705 (US). **YAMAMOTO, Stuart, Masakazu**; C/o Honda Patents & Technologies, North America, Llc, 21001 State Route 739, Raymond, OH 43067-9705 (US). **EL DOKOR, Tarek**; C/o Honda Patents & Technologies North, North America, Llc, 21001 State Route 739, Raymond, OH 43067-9705 (US). **KING, Josh**; C/o Honda Patents & Technologies, North America, Llc, 21001 State Route 739, Raymond, OH 43067-9705 (US).(74) Agents: **MCNELIS, John, T.** et al.; Fenwick & West LLP, Silicon Valley Center, 801 California Street, Mountain View, CA 94041 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: GESTURE AND VOICE RECOGNITION INTO A SINGLE SYSTEM

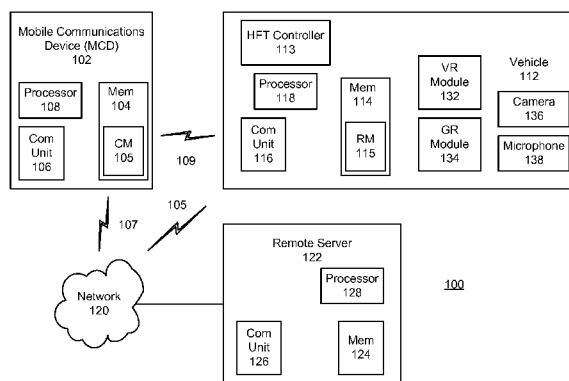


Figure 1

(57) Abstract: A system and method for combining two separate types of human machine interfaces, e.g., a voice signal and a gesture signal, performing voice recognition to a voice signal and gesture recognition to the gesture signal. Based on a confidence determination using the voice recognition result and the gesture recognition result the system can, for example, immediately perform the command/request, request confirmation of the command/request or determine that the command/request was not identified.

GESTURE AND VOICE RECOGNITION INTO A SINGLE SYSTEM

FIELD OF THE INVENTION

[0001] This application relates to gesture and voice recognition systems and more particularly to gesture and voice command recognition in a vehicle.

BACKGROUND

[0002] Human machine interfaces (HMIs) are common in many consumer devices and vehicles. A problem with conventional systems is that human machine interfaces are limited in the ability to recognize a user's command/request based on, for example a voice signal or visual signal. A more robust HMI and command/request recognition system is needed.

SUMMARY

[0003] A system and method for combining two separate types of human machine interfaces, e.g., a voice signal and a gesture signal, and performing voice recognition to a voice signal and gesture recognition to the gesture signal. Based on a confidence determination using the voice recognition result and the gesture recognition result the system can, for example, immediately perform the command/request, request confirmation of the command/request or determine that the command/request was not identified.

[0004] The features and advantages described in the specification are not all inclusive and, in particular, many additional features and advantages will be apparent to one of ordinary skill in the art in view of the drawings, specification, and claims. Moreover, it should be noted that the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter. For example, while the use in a vehicle is set forth herein as an example for ease of discussion, the present invention is not limited to use in a vehicle

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Figure 1 is an illustration of an environment in which one embodiment may operate.

- [0006] Figure 2 is an illustration of various embodiments for activating a voice recognition system and a gesture recognition system.
- [0007] Figure 3 is a flowchart of a method in accordance with one embodiment.
- [0008] Figure 4 is an illustration setting forth an example of the effect of a gesture recognition score and a voice recognition score when the commands are in the same domain in accordance with one embodiment.
- [0009] Figure 5 is an illustration setting forth an example of the effect of a gesture recognition score and a voice recognition score when the command with the highest voice recognition score differs from the command with the highest gesture recognition score in accordance with one embodiment.
- [0010] The figures depict various embodiments for purposes of illustration only. One skilled in the art will readily recognize from the following discussion that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles described herein.

DETAILED DESCRIPTION

- [0011] Embodiments are now described with reference to the figures where like reference numbers indicate identical or functionally similar elements. Also in the figures, the left most digit of each reference number corresponds to the figure in which the reference number is first used.
- [0012] Figure 1 illustrates an exemplary operating environment 100 for various embodiments. Operating environment 100 may include an in-vehicle communications system, one example of such a system is an in-vehicle hands free telephone (HFT) controller 113 which will be used as an example herein for ease of discussion. The operating environment 100 may also include a wireless mobile communication device (MCD) 102, a communication link 105 for communications between the in-vehicle system 112 and a network 120, a short-range communication link 109 for communication between the in-vehicle system 112 and wireless mobile communication device 102, a wireless networking communication link 107 between wireless mobile communication device 102 and network 120, and a processing device, such as a server 122 connected to network 120. The communication links described herein can directly or indirectly connect these devices. The network 120 can be a wireless communication network such as a cellular network comprised of multiple base stations, controllers, and

a core network that typically includes multiple switching entities and gateways, for example.

[0013] The vehicle 112 can include a camera 136, a microphone 138, a voice recognition (VR) module 132 and/or a gesture recognition (GR) module 134. The camera 136 and/or microphone can be in the vehicle 112 or can be in the MCD 102 (not shown), for example. In an embodiment the camera 136 can be any optical sensing device, for example, a sensor that captures signals from within the vehicle, e.g., a time of flight camera, an infrared sensor, a traditional camera, CMOS type sensor cameras, etc. The camera 136 is positioned to receive signals from a user such as hand signals from a driver or passenger. In various embodiments, the signals received by the camera 136 represent gestures by a driver/passenger. The camera 136 can include multiple cameras 136 positioned to receive signals from various positions in the vehicle, e.g., driver seat, front passenger seat, second row seats, etc. Alternatively a single camera can be used which is focused on one position, e.g., the driver, has a wide field of view and can receive signals from more than one occupant of the vehicle, or can change its field of view to receive signals from different occupant positions. In another embodiment, the camera 136 is positioned in the MCD 102, e.g., a camera incorporated into a smart phone, and the camera can be positioned to capture the gestures by the occupant. For example, the camera 136 can be mounted so that it faces the driver and can capture gestures by the driver. The camera 136 may be positioned in the cabin or pointing toward the cabin and can be mounted on the ceiling, headrest, dashboard or other locations in/on the vehicle 112 or MCD 102.

[0014] The vehicle can include a microphone 138 that receives audio signals from inside the vehicle. In an embodiment the microphone can be positioned such that it is more sensitive to sound emanating from a particular position, e.g., the position of the driver, than other positions, e.g., other occupants. The microphone 138 can be standard microphone that is incorporated into a vehicle or can be a microphone incorporated into a MCD 102. The microphone 138 can be mounted so that it receives voice signals from the driver. For example, the microphone 138 may be positioned in the cabin or pointing toward the cabin and can be mounted on the ceiling, headrest, dashboard or other locations in/on the vehicle 112 or MCD 102.

[0015] In the embodiment illustrated in Figure 1, the vehicle includes a voice recognition module 132 and a gesture recognition module 136. The voice recognition

module 132 can include software that receives an output signal of the microphone 138 and performs a voice recognition algorithm on the received signal to identify the word or words received by the microphone 138. In alternate embodiments, the voice recognition module 132 is positioned external to the vehicle 132, for example the voice recognition module can be positioned in the MCD 102 or a remote server 122. For example, the voice recognition module 132 can be an application downloaded to the MCD 102, e.g., an application available from iTunes. The gesture recognition module 134 can include software that receives an output signal of the camera 138 and performs a gesture recognition algorithm on the received signal to identify/interpret the gesture received by the camera 136. In alternate embodiments, the gesture recognition module 134 is positioned external to the vehicle 132, for example the gesture recognition module 134 can be positioned in the MCD 102 or a remote server 122. For example, the gesture recognition module 134 can be an application downloaded to the MCD 102, e.g., an application available from iTunes.

[0016] In-vehicle hands free telephone (HFT) controller 113 and wireless mobile communication device 102 may communicate with each other via a short-range communication link 109 which uses short-range communication technology, such as, for example, Bluetooth® technology or other short-range communication technology, for example, Universal Serial Bus (USB). In-vehicle HFT controller system 113 and wireless mobile communication device 102 may connect, or pair, with each other via short-range communication link 109. In an embodiment the HFT controller 113 can include a communications unit 116 to assist in the short-range communications, a memory unit device 114, and a processor 118. The HFT controller 113 can be part of a vehicle's telematics system, which includes memory/storage, processor(s) and communication unit(s). Figure 1 shows the memory 114, communication unit 116 and processor 118 as being part of the in-vehicle HFT controller 113 for ease of discussion. In alternate embodiments the HFT controller 113 can utilize the vehicle's telematics unit to assist in performing various functions. For example the communications unit 116 and/or processor 118, for example, can be part of the vehicle's telematics unit or can be a separate unit in the vehicle. The MCD 102 has an operating system and can include various applications either integrated into the operating system or stored in memory/storage 104 and executed by the processor 108.

[0017] Processors 108, 118 and/or 128 process data signals and may comprise various computing architectures including a complex instruction set computer (CISC) architecture, a reduced instruction set computer (RISC) architecture, or an architecture implementing a combination of instruction sets. Although only a single processor is shown in figure 1, multiple processors may be included. The processors can comprises an arithmetic logic unit, a microprocessor, a general purpose computer, or some other information appliance equipped to transmit, receive and process electronic data signals from the memory 104, 114, 124, and other devices both shown and not shown in the figures.

[0018] Examples of a wireless mobile communication device (MCD) 102 include a cellular phone, personal device assistant (PDA), smart phone, pocket personal computer (PC), laptop computer, tablet computer, smart watch or other devices having a processor, communications capability and are easily transportable, for example. In a common form, an MCD application can be part of a larger suite of vehicle features and interactions. Examples of applications include applications available for the iPhone™ that is commercially available from Apple Computer, Cupertino, California, applications for phones running the Android™ operating system that is commercially available from Google, Inc., Mountain View, California, applications for BlackBerry devices, available from RIM, Ontario Canada, and/or applications available for Windows Mobile devices, available from Microsoft Corp., Redmond, Washington. In an embodiment the MCD 102 includes a communications unit 106 a memory unit device 104, and a processor 108. The MCD 102 has an operating system and can include various applications either integrated into the operating system or stored in memory/storage 104 and executed by the processor 108.

[0019] In alternate embodiments a mobile communication device 102 can be used in conjunction with a communication device embedded in the vehicle, such as a vehicle embedded phone, a wireless network card or other device (e.g., a Wi-Fi capable device). For ease of discussion the description herein describes the operation of the embodiments with respect to an embodiment using a mobile communication device 102. However, this is not intended to limit the scope of the embodiments and it is envisioned that other embodiments operate using other communication systems between the in-vehicle system 112 and the network 120, examples of which are described herein.

[0020] In-vehicle system 112 may send information to wireless mobile communication device 102. Wireless mobile communication device 102 may send information to in-vehicle system 112 via short-range communication link 109. Wireless mobile communication device 102 may store information received from in-vehicle system 112, and/or may provide the information (such as voice and/or gesture signals) to a remote processing device, such as, for example, server 122, via network 120. Remote server 122 can include a communications unit 126 to connect to the network 120, for example, a memory/storage unit 124 and a processor 128.

[0021] In some embodiments, in-vehicle system 112 may provide information to the wireless mobile communication device 102. Wireless mobile communication device 102 may use that information to obtain additional information from network 120 and/or server 122. The additional information may also be obtained in response to providing information with respect to a prompt on wireless mobile communication device 102 from in-vehicle system 112.

[0022] Network 120 may include a wireless communication network, for example, a cellular telephony network, as well as one or more other networks, such as, the Internet, a public-switched telephone network (PSTN), a packet-switching network, a frame-relay network, a fiber-optic network, and/or other types of networks.

[0023] Figure 2 is an illustration of various embodiments for activating a voice recognition system and a gesture recognition system. A driver can interact with a gesture recognition module 134 and/or a voice recognition module 132 in a variety of ways.

[0024] In an embodiment, the user, e.g., a vehicle driver or passenger, interacts 202 with the gesture recognition (GR) module 134 independently of the voice recognition module 136. In this embodiment, the gesture recognition module 134 enables the driver to enter into different modes, such as, navigation, hands-free calling, music search, etc.

[0025] In another embodiment, the driver (or passenger) interacts 204 with the voice recognition (VR) module 136 independently of the gesture recognition module 134. In this embodiment the user may, for example, say a command which takes the voice recognition system into a particular mode or context, such as, navigation, web search, hands-free calling, music search, etc. The system may require the pressing of a talk switch or some other method of waking up/activating the voice recognition module 132

and/or microphone 138 (together the voice recognition system). In some embodiments, the voice recognition system supports an open microphone in which the system wakes-up when the user uses a specific word.

[0026] In another embodiment, the driver (or passenger) interacts 206 with the gesture recognition (GR) module 134 to trigger the voice recognition module 136. In one example of this embodiment the user may make a gesture that represents a phone, e.g., by gesturing the shape of a phone or a gesture that is interpreted as making a phone call, e.g., placing a hand near an ear with the thumb and pinky extended. The system takes the user to the phone mode and enables/opens the voice recognition system (VR module 132 and microphone 138) which allows the microphone 138 to receive and act upon a voice signal, e.g., a command, spoken by a user. In alternate embodiments, a voice prompt (informative or confirmative) may occur before enabling the voice recognition system.

[0027] In another embodiment, the driver (or passenger) interacts 208 with the voice recognition (VR) module 136 to trigger the gesture recognition module 134. In one example of this embodiment the user may speak a trigger word or phrase, e.g., “music.” The system can then go into a music mode and the user can scroll through album covers displayed in a vehicle’s display device by performing a gesture interpreted to mean “scroll.” In an alternate embodiment, a voice prompt (informative or confirmative) may occur before enabling the gesture recognition system.

[0028] In some embodiments the four methods of recognitions 202, 204, 206, 208 can alternate so that one method is active at a given time. In such a system, the user should follow the proper sequence to perform a desired task. In other embodiments, more than one method can be enabled simultaneously which provides additional flexibility to the user in having the desired task performed.

[0029] Figure 3 is a flowchart of a method in accordance with one embodiment. The embodiments can operate using any of the four methods of recognition 202, 204, 206, 208 described above. In this embodiment, a microphone 138 receives 302 a voice signal from a user in a vehicle. A camera 136 receives 304 a gesture signal. These two events can occur simultaneously or sequentially in any order.

[0030] The voice recognition module 132 performs a conventional voice recognition algorithm and in one embodiment generates 306 a voice recognition score. In an embodiment a voice recognition score can represent a confidence level that the output

of the voice recognition module 132 correctly matches the command represented by the voice command, e.g., a score of 90 can represent a 90 percent confidence value that the recognized phrase is correct.

[0031] The gesture recognition module 134 performs a gesture recognition algorithm and in one embodiment generates 308 a gesture recognition score. Any well-known gesture recognition algorithm may be employed. For example, one or more Markov models for dynamic gestures, Support Vector Machines for static gestures, and the like may be employed. Indeed any gesture recognition algorithm that interprets input gestures and provides a recognition score indicative of a level of confidence may be employed. In an embodiment a gesture recognition score can represent a confidence level that the output of the gesture recognition module 134 correctly matches the command represented by the gesture signal, e.g., a score of 90 can represent a 90 percent confidence value that the recognized gesture is correct. Steps 306 and 308 can be performed simultaneously or sequentially in any order.

[0032] In an embodiment, the gesture score and voice scores may be weighted. Alternatively, the system and method may use trained action units that are capable of defining the relationship itself. Ideally, a learning system will choose between the two modalities, or choose combining them, as well as combining other modalities such as touch. Additional details about techniques involving such recognition are known, e.g., in Pattern Classification, 2nd edition by Duda, Hart and Stork, which is incorporated by reference herein in its entirety.

[0033] In one embodiment, the recognition module 115 receives the voice recognition score and the gesture recognition score for the top one or more recognized strings, phrases, etc, and, in one embodiment, determines a combined recognition score based on the voice recognition score and the gesture recognition score for the top one or more recognized phrases. The phrase/command associated with combined recognitions score(s) is the “identified command.”

[0034] In one embodiment, the recognition module 115 receives the voice recognition score and the gesture recognition score for the top one or more recognized phrases and, in one embodiment, determines the phrase/command with the highest combined score (the “identified command”). As described below, in alternate embodiments, scores related to different commands/phrases can be identified. For the identified command the system determines 310 whether the combination of the scores exceeds a first

threshold (T1). If 310 the combination exceeds a first threshold then it is determined that the recognized phrase/command is correct and the phrase/command is executed 312. For example, in an embodiment the voice recognition score and the gesture recognition score are combined, such as a summing of the two scores, although many other arithmetic factors/equations can be used, and if the total exceeds the first threshold then it's determined that the phrase/command is correct and the command is executed 312.

[0035] If 310 the system determines that the combination of the voice recognition score and gesture recognition score for the phrase/command does not exceed a first threshold but is above 314 a second threshold (T2) then the system requests confirmation 316 from the user. The confirmation can be in various formats, such as one or more of a verbal signal, a visual signal, and/or a tactile signal, for example. The confirmation may require the user to actively confirm that the identified phrase/command is correct or can provide the user with the identified phrase/command and execute the command unless the user indicates that the phrase/command is incorrect.

[0036] If 314 the system determines that the combination of the voice recognition score and gesture recognition score for the phrase/command does not exceed the second threshold then the system determines that the identified command is not recognized 320 at a high enough confidence level to proceed. In an embodiment the system requests that the user to repeat the command.

[0037] Figure 4 is an illustration setting forth an example of the effect of a gesture recognition score and a voice recognition score when the commands are in the same domain in accordance with one embodiment. In the example set forth in Figure 4, four cases are set forth based on a values of the gesture recognition score and voice recognition score. This is another embodiment that performs steps similar to steps 310-316 in Figure 3. In Case 1, if the gesture recognition score exceeds a threshold (T3) and the voice recognition score exceeds a threshold (T4) then the system executes the command.

[0038] In Case 2, if the gesture recognition score does not exceed threshold T3 but the voice recognition score exceeds threshold T4 then the system requests confirmation of the identified phrase/command from the user.

[0039] In Case 3, if the gesture recognition score exceeds threshold T3 but the voice recognition score does not exceed threshold T4 then the system requests confirmation of the identified phrase/command from the user.

[0040] In Case 4, if the gesture recognition score does not exceed threshold T3 and the voice recognition score does not exceed threshold T4 then the system determines that the phrase/command was not recognized. In an embodiment the system requests that the user repeat the command.

[0041] In embodiments, the threshold values T3 and T4 can be the same, while in other embodiments they are different. In addition, additional thresholds, e.g., T5, T6 etc. can be used when determining whether the phrase/command is properly identified based upon the voice recognition score and the gesture recognition score. For example, additional threshold comparisons may be done based on different thresholds, e.g., T5, T6 so that in Case 2, if the gesture recognition score is below T3 but is above T5 then the confirmation requested is one where the vehicle will perform the command unless the user indicates that the identified command is incorrect. However if the gesture recognition score is also below T3 and T5 (with the Voice recognition score still being above T4) then the vehicle may request that the user affirmatively confirm the identified phrase/command before the command is executed. Examples of such a confirmation include pressing a button, performing a gesture, speaking or a combination thereof.

[0042] In alternate embodiments, if the gesture recognition score is below a threshold, e.g., threshold T6, or the voice recognition score below a threshold, e.g., threshold T6 or a different threshold T7, then the system will determine that the command is not recognized.

[0043] Figure 5 is an illustration setting forth an example of the effect of a gesture recognition score and a voice recognition score when the command with the highest voice recognition score differs from the command with the highest gesture recognition score in accordance with one embodiment. Note that the threshold values (T3, T4) are not necessarily the same as those described above with respect to Figure 4. In the examples set forth above the phrase/command with the best combined score was compared. Alternatively the best voice recognition score and the best gesture recognition score were both associated with the same phrase/command. In an alternate

embodiment, phrase/command receiving the highest voice recognition command are different.

[0044] In Case 1, if the gesture recognition score for phase/command “X” is above threshold T3 and the voice recognition score for phrase/command “Y” is above threshold T4 (which may be the same or different than T3) then the system queries the user regarding whether the user wants to perform command X or Y.

[0045] In Case 2, if the gesture recognition score for phase/command “X” is below threshold T3 and the voice recognition score for phrase/command “Y” is above threshold T4 then the system queries the user to confirm that the user wants to perform command Y.

[0046] In Case 3, if the gesture recognition score for phase/command “X” is above threshold T3 and the voice recognition score for phrase/command “Y” is below threshold T4 then the system queries the user to confirm that the user wants to perform command X.

[0047] In Case 4, if the gesture recognition score for phase/command “X” is below threshold T3 and the voice recognition score for phrase/command “Y” is below threshold T4 then the system determines that it has not found a match and may ask the user to try again.

[0048] Reference in the specification to “one embodiment” or to “an embodiment” means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least one embodiment. The appearances of the phrase “in one embodiment” or “an embodiment” in various places in the specification are not necessarily all referring to the same embodiment.

[0049] Some portions of the detailed description are presented in terms of algorithms and symbolic representations of operations on data bits within a computer memory. These algorithmic descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps (instructions) leading to a desired result. The steps are those requiring physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical, magnetic or optical signals capable of being stored, transferred, combined, compared and otherwise manipulated. It is convenient at times, principally for reasons of common usage, to refer to these

signals as bits, values, elements, symbols, characters, terms, numbers, or the like. Furthermore, it is also convenient at times, to refer to certain arrangements of steps requiring physical manipulations or transformation of physical quantities or representations of physical quantities as modules or code devices, without loss of generality.

[0050] However, all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the following discussion, it is appreciated that throughout the description, discussions utilizing terms such as “processing” or “computing” or “calculating” or “determining” or “displaying” or “determining” or the like, refer to the action and processes of a computer system, or similar electronic computing device (such as a specific computing machine), that manipulates and transforms data represented as physical (electronic) quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0051] Certain aspects of the embodiments include process steps and instructions described herein in the form of an algorithm. It should be noted that the process steps and instructions of the embodiments can be embodied in software, firmware or hardware, and when embodied in software, could be downloaded to reside on and be operated from different platforms used by a variety of operating systems. The embodiments can also be in a computer program product, which can be executed on a computing system.

[0052] The embodiments also relate to an apparatus for performing the operations herein. This apparatus may be specially constructed for the purposes, e.g., a specific computer, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, application specific integrated circuits (ASICs), or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus. Memory can include any of the above and/or other devices that can store information/data/programs and can be transient or non-transient medium,

where a non-transient or non-transitory medium can include memory/storage that stores information for more than a minimal duration. Furthermore, the computers referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[0053] The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general-purpose systems may also be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the method steps. The structure for a variety of these systems will appear from the description herein. In addition, the embodiments are not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the embodiments as described herein, and any references herein to specific languages are provided for disclosure of enablement and best mode.

[0054] In addition, the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure of the embodiments is intended to be illustrative, but not limiting, of the scope of the embodiments, which is set forth in the claims.

[0055] While particular embodiments and applications have been illustrated and described herein, it is to be understood that the embodiments are not limited to the precise construction and components disclosed herein and that various modifications, changes, and variations may be made in the arrangement, operation, and details of the methods and apparatuses of the embodiments without departing from the spirit and scope of the embodiments as defined in the appended claims.

What is claimed is:

1. A computer based method for executing a command in a vehicle, comprising the steps of:
receiving a voice signal;
receiving a gesture signal;
performing a voice recognition procedure to generate a voice recognition (VR) score representing a likelihood that the voice signal corresponds to a first voice command;
performing a gesture recognition procedure to generate a gesture recognition (GR) score representing a likelihood that the gesture signal corresponds to a first gesture command; and
executing a first command in a vehicle corresponding to said first voice command and first gesture command when said voice recognition score exceeds a first threshold and said gesture recognition score exceeds a second threshold.
2. The method of claim 1, further comprising the step of:
generating a confirmation request, prior to executing said first command, when either (a) said voice recognition score exceeds a first threshold and said gesture recognition score does not exceed a second threshold or (b) said voice recognition score does not exceed a first threshold and said gesture recognition score exceeds a second threshold.
3. The method of claim 2, further comprising the step of:
not executing a command corresponding to said first voice command or said first gesture command when said voice recognition score does not exceed a first threshold and said gesture recognition score does not exceed a second threshold.
4. The method of claim 1, wherein said confirmation request is at least one of an audible request and a visual request.

5. The method of claim 1, further comprising the step of:
not executing a command corresponding to said first voice command or said first gesture command when said voice recognition score does not exceed a first threshold and said gesture recognition score does not exceed a second threshold.
6. A computer based method for executing a command in a vehicle, comprising the steps of:
receiving a voice signal;
receiving a gesture signal;
performing a voice recognition procedure to generate a voice recognition (VR) score representing a likelihood that the voice signal corresponds to a first voice command;
performing a gesture recognition procedure to generate a gesture recognition (GR) score representing a likelihood that the gesture signal corresponds to a first gesture command;
generating a combined recognition score based on a combination of said VR score and said GR score; and
executing a first command in the vehicle corresponding to at least one of said first voice command and said first gesture command when said combined recognition score exceed a first threshold.
7. The method of claim 6, further comprising the step of:
generating a confirmation request, prior to executing said first command, when said combined recognition score does not exceed a second threshold.
8. The method of claim 7, wherein said first and second thresholds are substantially the same.
9. The method of claim 6, wherein said confirmation request is at least one of an audible request and a visual request.

10. The method of claim 6, wherein said voice signal and gesture signal are received in a vehicle.
11. The method of claim 6, wherein said step of performing said gesture recognition procedure is performed in a mobile communication device.
12. The method of claim 11, wherein said mobile communication device is one of a cellular phone, a personal device assistant, a smart phone, a pocket personal computer, a laptop computer, a tablet computer, or a smart watch.
13. A non-transitory computer readable medium comprising computer executable code which when executed by a computer performs the steps of:
 - receiving a voice signal;
 - receiving a gesture signal;
 - performing a voice recognition procedure to generate a voice recognition (VR) score representing a likelihood that the voice signal corresponds to a first voice command;
 - performing a gesture recognition procedure to generate a gesture recognition (GR) score representing a likelihood that the gesture signal corresponds to a first gesture command;
 - executing a first command in a vehicle corresponding to said first voice command and first gesture command when said voice recognition score exceeds a first threshold and said gesture recognition score exceeds a second threshold.
14. The computer readable medium of claim 13, further comprising the step of:
 - generating a confirmation request, prior to executing said first command, when either (a) said voice recognition score exceeds a first threshold and said gesture recognition score does not exceed a second threshold or (b) said voice recognition score does not exceed a first threshold and said gesture recognition score exceeds a second threshold.

15. The computer readable medium of claim 14, further comprising the step of:
not executing a command corresponding to said first voice command or said
first gesture command when said voice recognition score does not
exceed a first threshold and said gesture recognition score does not
exceed a second threshold.
16. The computer readable medium of claim 13, wherein said confirmation
request is at least one of an audible request and a visual request.
17. The computer readable medium of claim 13, further comprising the step of:
not executing a command corresponding to said first voice command or said
first gesture command when said voice recognition score does not
exceed a first threshold and said gesture recognition score does not
exceed a second threshold.
18. A non-transitory computer readable medium comprising computer executable
code which when executed by a computer performs the steps of:
receiving a voice signal;
receiving a gesture signal;
performing a voice recognition procedure to generate a voice recognition (VR)
score representing a likelihood that the voice signal corresponds to a
first voice command;
performing a gesture recognition procedure to generate a gesture recognition
(GR) score representing a likelihood that the gesture signal
corresponds to a first gesture command;
generating a combined recognition score based on a combination of said VR
score and said GR score; and
executing a first command in a vehicle corresponding to at least one of said
first voice command and said first gesture command when said
combined recognition score exceed a first threshold.
19. The computer readable medium of claim 18, further comprising the step of:

generating a confirmation request, prior to executing said first command, when said combined recognition score does not exceed a second threshold.

20. The computer readable medium of claim 18, wherein said confirmation request is at least one of an audible request and a visual request.
21. A system for executing a command in a vehicle comprising:
 - a microphone positioned in the vehicle, for receiving a voice signal;
 - a optical sensing device, positioned in the vehicle, for receiving a gesture signal;
 - a processor for performing a voice recognition procedure to generate a voice recognition (VR) score representing a likelihood that the voice signal corresponds to a first voice command, the processor for performing a gesture recognition procedure to generate a gesture recognition (GR) score representing a likelihood that the gesture signal corresponds to a first gesture command; and
 - an execution unit positioned in the vehicle, for executing a first command corresponding to said first voice command and first gesture command when said voice recognition score exceeds a first threshold and said gesture recognition score exceeds a second threshold.
22. The system of claim 21, further comprising:
 - confirmation unit, for generating a confirmation request, prior to executing said first command, when either (a) said voice recognition score exceeds a first threshold and said gesture recognition score does not exceed a second threshold or (b) said voice recognition score does not exceed a first threshold and said gesture recognition score exceeds a second threshold.

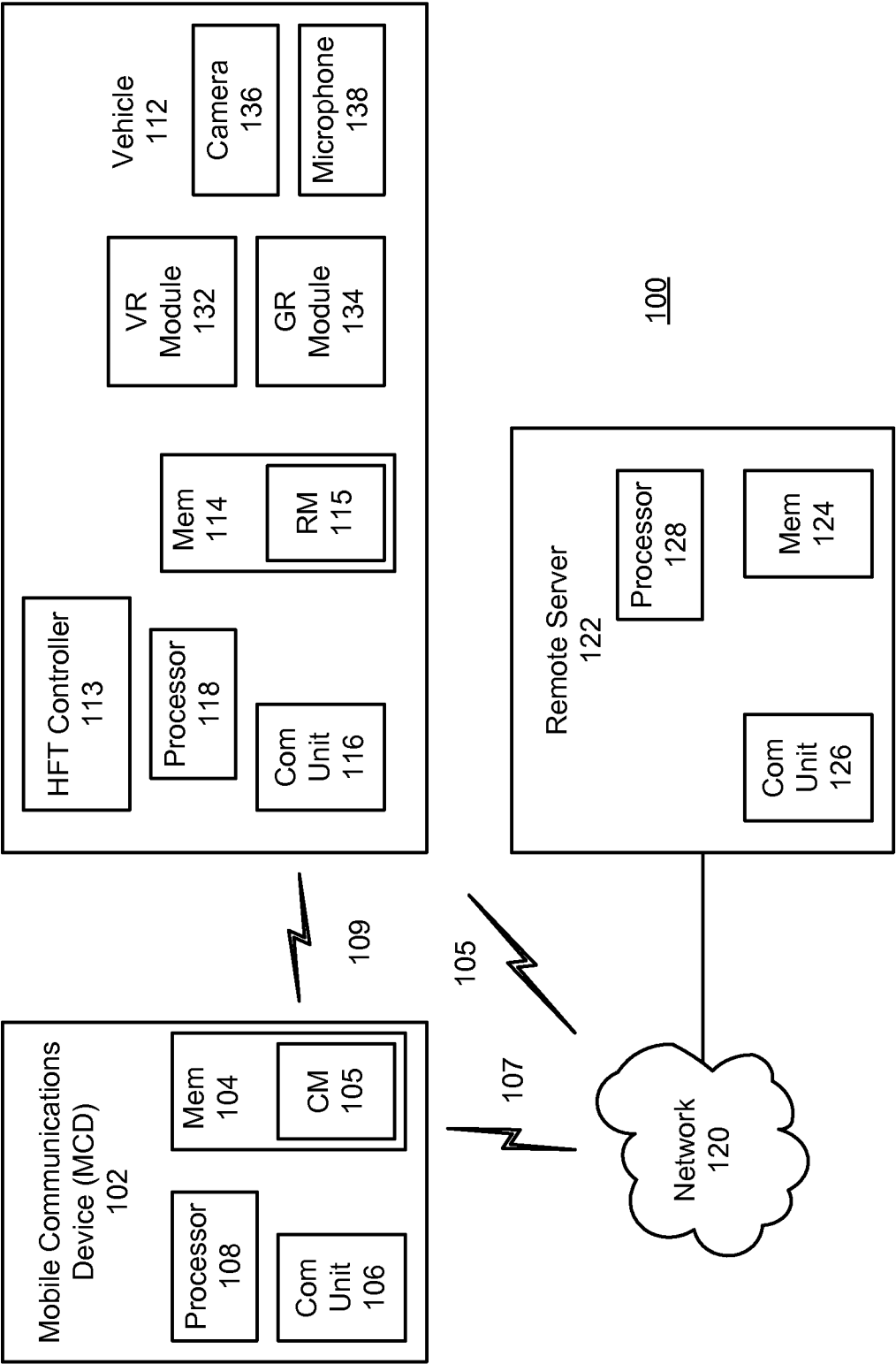


Figure 1

2/5

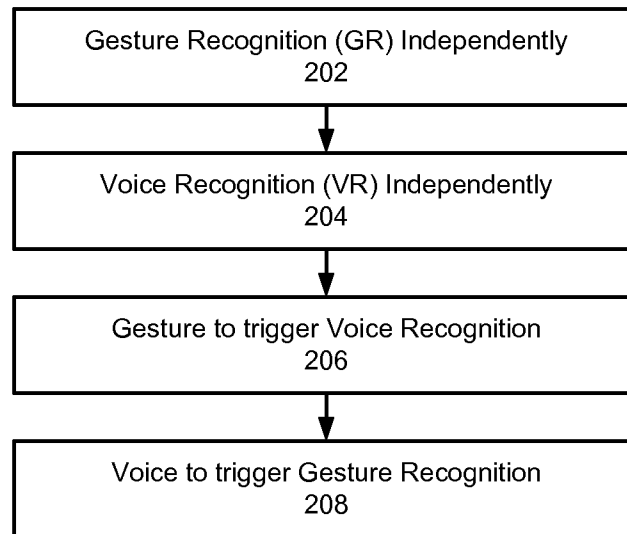


Figure 2

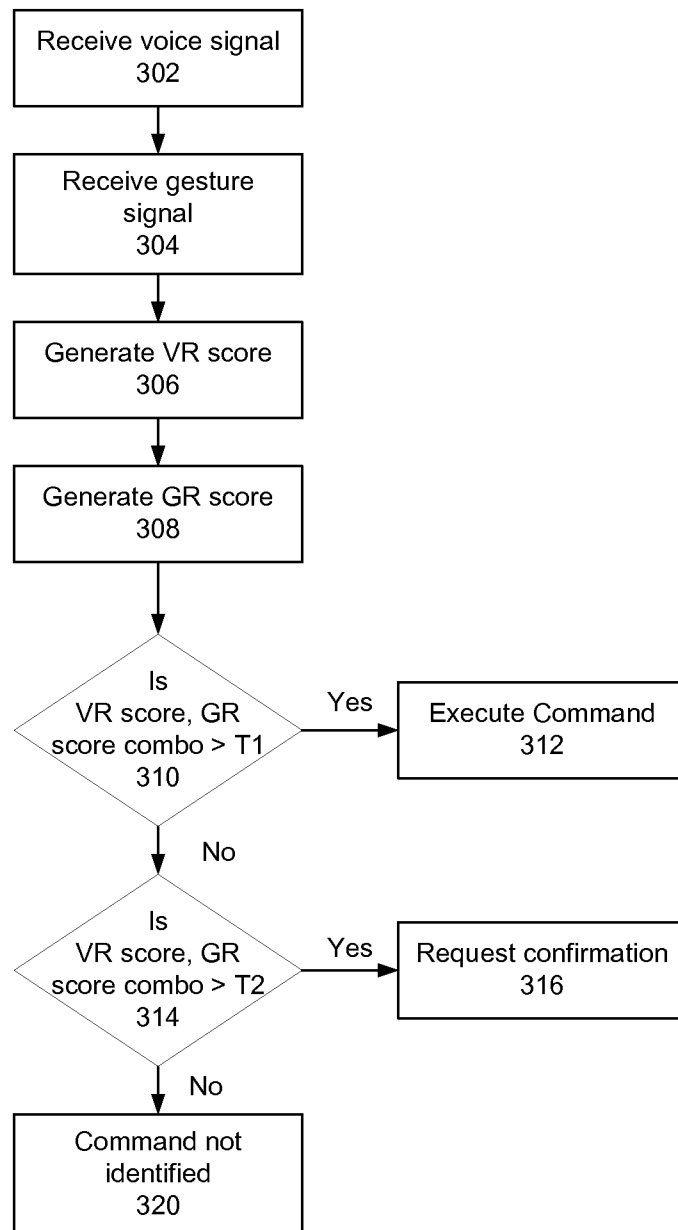


Figure 3

	GR Score	VR Score	Decision
Case 1	$\geq T3$	$\geq T4$	Execute command
Case 2	$< T3$	$\geq T4$	Confirmation requested
Case 3	$\geq T3$	$< T4$	Confirmation requested
Case 4	$< T3$	$< T4$	Command not recognized

Figure 4

Commands X and Y Differ

	GR Score Command X	VR Score Command Y	Decision
Case 1	$\geq T3$	$\geq T4$	Ask user, Do you want to perform X or Y
Case 2	$< T3$	$\geq T4$	Ask user for confirmation to perform Y
Case 3	$\geq T3$	$< T4$	Ask user for confirmation to perform X
Case 4	$< T3$	$< T4$	Command not recognized

Figure 5

INTERNATIONAL SEARCH REPORT		International application No. PCT/US14/13070		
A. CLASSIFICATION OF SUBJECT MATTER IPC: G06F 17/00 (2006.01) USPC: 701/001 According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 701/001 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) combined, score, voice, gesture, value, signal, sensor, first, second, recognition				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
A	US 8,095,291 B2 (CHRISTEN ET AL) 10 January 2012 (10.01.2012), see entire document.	1-22		
A	US 6,792,339, B2 (BASSON ET AL) 14 September 2004 (14.09.2004), see entire document.	1-22		
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.				
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top;"> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="width: 50%; vertical-align: top;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search 28 May 2014 (28.05.2014)		Date of mailing of the international search report 23 JUN 2014		
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201		Authorized officer Ken Wieder Telephone No. 571-272-2600		