(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 June 2014 (26.06.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/096506 A1

(51) International Patent Classification:
G10L 15/065 (2013.01) *G10L 15/30* (2013.01)

(21) International Application Number:
PCT/FI2012/051285

(22) International Filing Date:
21 December 2012 (21.12.2012)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: NOKIA CORPORATION [FI/FI]; Keilalahdentie 4, FI-02150 Espoo (FI).

(72) Inventor: PAK, Yongbeom; Ylioppilaantie 8 B 20, FI-90130 Oulu (FI).

(74) Agents: NOKIA CORPORATION et al.; IPR Department, Jussi Jaatinen, Keilalahdentie 4, FI-02150 Espoo (FI).

(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, IS, JP, KE, KG, KM, 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, 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, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

[Continued on next page]

(54) Title: METHOD, APPARATUS, AND COMPUTER PROGRAM PRODUCT FOR PERSONALIZING SPEECH RECOGNITION

(57) Abstract: A method, apparatus and computer program product are provided for personalizing speech recognition data. A speech recognition model (SRM) that is adaptable by a user terminal based on user terminal dependent data may be received and adapted by a user terminal. A speaker dependent SRM may be refined on the user terminal and transmitted to a remote storage location, such as personalized speech recognition apparatus. The apparatus may cause transmission of SRMs to various user terminals, and may generate additional SRMs based on speaker dependent SRMs. Speaker dependent SRMs may be generated based on an individual, group of users, geographic location, dialect, or the like. SRMs may be based on hidden Markov Models,dynamic time warming models, neural networks, finite state transducers, or the like.

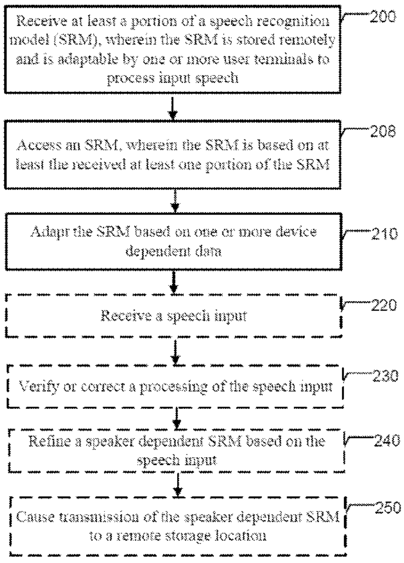
Receive at least a portion of a speech recognition model (SRM), wherein the SRM is stored remotely and is adaptable by one or more user terminals to process input speech
200
Access an SRM, wherein the SRM is based on at least the received at least one portion of the SRM
208
Adapt the SRM based on one or more device dependent data
210
Receive a speech input
220
Verify or correct a processing of the speech input
230
Refine a speaker dependent SRM based on the speech input
240
Cause transmission of the speaker dependent SRM to a remote storage location
250

Figure 2

WO 2014/096506 A1

Published:

— *with international search report (Art. 21(3))*

METHOD, APPARATUS, AND COMPUTER PROGRAM PRODUCT FOR PERSONALIZING SPEECH RECOGNITION

TECHNOLOGICAL FIELD

An example embodiment of the present invention relates generally to speech recognition, and more particularly, to a method, apparatus and computer program product for personalizing speech recognition.

BACKGROUND

The widespread use of technology, including mobile technology, in everyday life has led to an increased demand for other forms of user interaction with various devices. Devices providing a user with hands free control capabilities are becoming increasingly popular that allow users to control a device with voice commands, such as via speech recognition, while still focusing their attention on driving or other activities. Speech recognition may be used to control these and other devices, such as wireless phones, cars, household appliances, and other devices used in everyday life or work.

Speech recognition, which may be referred to as automatic speech recognition (ASR), may be conducted by various applications that may be operable to convert recognized speech into text (e.g., a speech-to-text system). Current ASR and/or speech-to-text systems are typically based on a speech recognition model (SRM) comprising an acoustic model and a language model. For improved efficiency, the acoustic modes and language models can be fused together, or otherwise may be combined. These SRMs are the building blocks for words and strings of words, such as phrases or sentences and are used by a device to process speech input (e.g., recognize the speech input and derive a machine readable interpretation).

By way of example, a speech recognition processor, in some examples, may receive speech samples and then may match those samples with the basic sound units in the acoustic model. The speech recognition processor then may, for example, calculate the most likely words from the SRM based on the matched basic sound units, such as by using Hidden Markov Models (HMMs) and/or dynamic time warping (DTW). HMM and DTW are examples of statistical models that describe speech patterns probabilistically. Additionally or alternatively, various neural networks (NN) and /or finite state transducers (FST) may also be used as SRMs. Other suitable models can also be used as SRM.

In the DTW and in some additional examples, an unknown speech pattern is compared with known reference patterns. In dynamic time warping, the speech pattern is divided into several frames, and the local distance between the speech pattern included in each frame and the corresponding speech segment of the reference pattern is calculated. This distance is calculated by comparing the speech segment and the corresponding speech segment of the reference pattern with each other, and it is thus a kind of numerical value for the differences found in the comparison. For speech segments close to each other, a smaller distance is usually obtained than for speech segments further from each other. On the basis of local distances obtained this way, a minimum path between the beginning and end points of the word are sought by using a DTW algorithm. Thus, by DTW, a distance is obtained between the uttered word and the reference word.

In speech recognition using the HMM method, an HMM model is first formed for each word to be recognized (e.g. for each reference word). When the speech recognition device receives a speech pattern, an observation probability is calculated for each HMM model in the memory, and as the recognition result, a counterpart word is obtained for the HMM model with the greatest observation probability. Thus for each reference word, the probability is calculated that it is the word uttered by the speaker. The above-mentioned observation probability describes the resemblance of the received speech pattern and the closest HMM model (e.g. the closest reference speech pattern). The reference words, or word candidates, can be further weighted by the language models. In some embodiments, the recognition process can occur in a single pass-through mode with fused acoustic models and language models.

In a NN method, interconnecting data nodes store information regarding speech patterns. The nodes of the NN may be used to classify phonetic features of speech input, and may be configured so as to focus on portions of the model that may be most valuable in distinguishing words during speech recognition processes. A well designed NN will therefore minimize, in some examples, the processing time required to recognize speech inputs. NNs are particularly well suited for training of larger data sets, such as data sets representing natural language.

In an FST method, speech inputs may be processed, various operations may be performed on the speech input, and a most probable output, (e.g., recognized word) may be selected. FSTs may be particularly beneficial, in some examples, in phonological analysis. The reusability and flexibility of algorithms performed on FSTs make FSTs particularly useful in combining portions of, or various SRMs. An SRM may therefore incorporate speech recognition data from various sources, apply weights to the speech recognition data, and generate weighted FSTs for use in speech recognition tasks.

The various types of SRMs may include speaker independent SRMs and speaker dependent SRMs. Speaker independent SRMs may comprise averages of language and acoustic models collected from a large sample of users. A speaker dependent SRM may be specific to the user and may be adapted by the user through training. Initial training may be performed during a first use of the SRM and training continues during normal use of the SRM. A speaker dependent SRM comprises unique sets of electronic characteristics for the acoustic model and a unique language model for the words formed from combinations of unique basic sound units.

It is appreciated that given the complexity of natural language, the data needed to process and understand speech may also be complex. SRMs used by a device to process speech input may therefore rely on any combination of the HMM, DTW, NN, FST, and other models, as well as a blend of speaker dependent SRMs and speaker independent SRMs.

BRIEF SUMMARY

A method, apparatus, and computer program product are provided for personalizing a speech recognition model (SRM). In one embodiment, a method is provided for receiving at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely and is adaptable by one or more user terminal to process input speech, accessing a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model, and adapting the speech recognition model based on terminal dependent data.

In some embodiments, the method may further include processing received speech input using the speech recognition model, and generating a textual output. In some embodiments, the method may further include receiving a speech input, and refining a speaker dependent speech recognition model based on the speech input. In some embodiments, the method may further include verifying or correcting a processing of the speech input, wherein refining the speaker dependent speech recognition model is further based on the verification or correction. In some embodiments, the method may further include causing transmission of at least one portion of the speaker dependent speech recognition model to a remote storage location. The terminal dependent data may comprise microphone information and/or a context. The received at least one portion of a speech recognition model is received based on one of at least an individual user, group of users, geographic location, or a dialect. The received at least one portion of a speech recognition model is based on at least one of a Hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

An additional method is provided including receiving at least one portion of a speaker dependent speech recognition model from a user terminal and generating at least one additional portion of a speech recognition model based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of a speech recognition model is adaptable by one or more user terminals.

In some embodiments, the method may further include causing transmission of the at least one additional portion of the speech recognition model to an additional user terminal. Generating the at least one additional portion of the speech recognition model is further based on one of at least an individual user, group of users, geographic location, or a dialect. The at least one additional portion of the speech recognition model may be based on at least one of a hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

An apparatus is also provided, comprising at least one processor and at least one memory including computer program code, the at least one memory and the computer program code configured to, with the processor, cause the apparatus to at least receive at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely, and is adaptable by one or more user terminals to process input speech, access a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model, and adapt the speech recognition model based on terminal dependent data.

An additional apparatus is provided comprising at least one processor and at least one memory including computer program code, the at least one memory and the computer program code configured to, with the processor, cause the apparatus to at least receive at least one portion of a speaker dependent speech recognition model from a user terminal, and generate at least one additional portion of a speech recognition model based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of the speech recognition model is adaptable by one or more user terminals.

A computer program product is provided, comprising at least one non-transitory computer-readable storage medium having computer-executable program code instructions stored therein, the computer-executable program code instructions comprising program code instructions to receive at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely, and is adaptable by one or more user terminals to process input speech, access a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model, and adapt the speech recognition model based on terminal dependent data.

An additional computer program product is provided, comprising at least one non-transitory computer-readable storage medium having computer-executable program code instructions stored therein, the computer-executable program code instructions comprising program code instructions to receive at least one portion of a speaker dependent speech recognition model from a user terminal, generate at least one additional portion of a speech recognition model based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of a speech recognition model is adaptable by one or more user terminals.

An apparatus is also provided, comprising means for receiving at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely, and is adaptable by one or more user terminals to process input speech, accessing a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model, and adapting the speech recognition model based on terminal dependent data.

An additional apparatus is provided, comprising means for receiving at least one portion of a speaker dependent speech recognition model from a user terminal, and generating at least one additional portion of a speech recognition based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of the speech recognition model is adaptable by one or more user terminals.

BRIEF DESCRIPTION OF THE DRAWINGS

Having thus described certain example embodiments of the present invention in general terms, reference will hereinafter be made to the accompanying drawings which are not necessarily drawn to scale, and wherein:

Figure 1 is a block diagram of a personalized speech recognition apparatus in communication with user terminals which may be configured to implement example embodiments of the present invention;

Figure 2 is a flowchart illustrating operations to receive and adapt an SRM on a user terminal, in accordance with one embodiment of the present invention;

Figure 3 is a flowchart illustrating operations to transmit an SRM, receive a speaker dependent SRM, and generate an additional SRM, using a speech personalization apparatus in accordance with one embodiment of the present invention; and

Figure 4 is a display for training an SRM, in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION

Some embodiments of the present invention will now be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all, embodiments of the invention are shown. Indeed, various embodiments of the invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Like reference numerals refer to like elements throughout. As used herein, the terms “data,” “content,” “information,” and similar terms may be used interchangeably to refer to data capable of being transmitted, received and/or stored in accordance with embodiments of the present invention. Thus, use of any such terms should not be taken to limit the spirit and scope of embodiments of the present invention.

Additionally, as used herein, the term ‘circuitry’ refers to (a) hardware-only circuit implementations (e.g., implementations in analog circuitry and/or digital circuitry); (b) combinations of circuits and computer program product(s) comprising software and/or firmware instructions stored on one or more computer readable memories that work together to cause an apparatus to perform one or more functions described herein; and (c) circuits, such as, for example, a microprocessor(s) or a portion of a microprocessor(s), that require software or firmware for operation even if the software or firmware is not physically present. This definition of ‘circuitry’ applies to all uses of this term herein, including in any claims. As a further example, as used herein, the term ‘circuitry’ also includes an implementation comprising one or more processors and/or portion(s) thereof and accompanying software and/or firmware. As another example, the term ‘circuitry’ as used herein also includes, for example, a baseband integrated circuit or applications processor integrated circuit for a mobile phone or a similar integrated circuit in a server, a cellular network device, other network device, and/or other computing device.

As defined herein, a “computer-readable storage medium,” which refers to a physical storage medium (e.g., volatile or non-volatile memory device), may be differentiated from a “computer-readable transmission medium,” which refers to an electromagnetic signal.

As described below, a method, apparatus and computer program product are provided for accessing and adapting remotely stored personalized speech recognition data for use on or more devices. Referring to Figure 1, personalized speech recognition apparatus 102 may include or otherwise be in communication with processor 20, user interface 22, communication interface 24, memory device 26, and speech personalization administrator 28. Personalized speech recognition apparatus 102 may be embodied by a wide variety of devices including mobile terminals, e.g., mobile telephones, smartphones, tablet computers, laptop computers, or the like, computers, workstations, servers or the like and may be implemented as a distributed system or a cloud based entity.

In example embodiments, the personalized speech recognition apparatus 102 may receive, and/or transmit SRMs, as well as generate additional SRMs that may be adaptable by one more user terminals. An SRM is a statistical model that describes speech patterns probabilistically, and may include a language model (words) and an acoustic model (basic sound units). Example

SRMs include the HMM, DTW, NN, and FST models. An SRM may be provided to a user terminal to enable speech recognition capabilities (e.g., processing of input speech) on the user terminal. In some embodiments, transmittal of an SRM may include transmittal of a portion of the SRM, since an SRM in its entirety may be too large for practical transmission (and an SRM portion may also be considered an SRM). The SRM portion may be incorporable into an SRM, so that the portion may then be incorporated with another portion of an SRM to provide a complete or fully functioning SRM. It will therefore be appreciated that any reference to an SRM herein, may indicate a portion or portions of an SRM, but for simplicity may be referred to as an SRM.

In some embodiments, the SRMs may incorporate speaker independent data, speaker dependent data, and/or terminal dependent data. The speaker independent data may include averaged, normalized, or otherwise consolidated language and acoustic models collected from a large sample of users.

The speaker dependent data may alternatively be biased toward a particular individual, or group of users, such as a group of users speaking a particular language or dialect, or from a particular geographic region. The speaker dependent data may be generated and/or refined on a user terminal by training the SRM. Alternatively or additionally, the speaker dependent data may be generated or refined on one or more user terminals and/or devices, such that it may be shared, via the personalized speech recognition apparatus 102, between the one or more user terminals and/or devices.

In some example embodiments, training may include, but is not limited to, providing speech input to the user terminal, potentially updating and/or verifying the processing of the speech input, and updating the SRM accordingly. On some user terminals, the training may include the explicit dictation of special training data by a speaker, and/or implicit training through the general use of the user terminal.

In some embodiments, various models, such as a HMM may be constructed for each the speaker dependent SRM to be stored. A speaker dependent SRM incorporating the speaker dependent data may be communicated from the user terminal to the personalized speech recognition apparatus 102.

The terminal dependent data may include information regarding the user terminal itself, such as characteristics of the microphone on the user terminal to capture the speech input, and/or a context of the user terminal (e.g., an environment the device is commonly used in, or the intended purpose of the device), or any settings of the user terminal 110A that could impact the processing of speech input. An SRM received from the personalized speech recognition apparatus 102 may be adapted on the user terminal based on the terminal dependent data, so that the particular user terminal may more accurately process speech inputs.

Speaker dependent SRMs, including speaker dependent data may be stored on personalized speech recognition apparatus 102. The speaker dependent SRM, or a portion thereof, may be further modified and/or transmitted to another device to allow the user terminal to benefit from the speaker dependent data, thereby improving the probability of successful speech recognition on another user terminal. As such, one or more user terminals may access or otherwise download the speaker dependent model for the purposes of providing personalized speech recognition.

Advantageously, for example, as the one or more user terminals provide personalized speech recognition, using the speaker dependent model, and the speech recognition result is verified or otherwise confirmed (e.g. check by a user for errors), the personalized speech recognition apparatus 102 may receive updates to the speaker dependent model. The personalized speech recognition apparatus 102 may therefore further tune or otherwise modify the speaker dependent model.

In some embodiments, the processor 20 (and/or co-processors or any other processing circuitry assisting or otherwise associated with the processor 20) may be in communication with the memory device 26 via a bus for passing information among components of the personalized speech recognition apparatus 102. The memory device 26 may include, for example, one or more volatile and/or non-volatile memories. In other words, for example, the memory device 26 may be an electronic storage device (e.g., a computer readable storage medium) comprising gates configured to store data (e.g., bits) that may be retrievable by a machine (e.g., a computing device like the processor 20). The memory device 26 may be configured to store information, data, content, applications, instructions, or the like for enabling the apparatus to carry out various functions in accordance with an example embodiment of the present invention. For example, the memory device 26 could be configured to store various SRMs, including speaker independent and speaker dependent portions. The speaker dependent data may be associated with a particular user or group of users, enabling the processor 20 to identify and provide appropriate SRMs to various devices. As such, the memory device 26 could be configured to buffer input data for processing by the processor 20, and/or to store instructions for execution by the processor 20.

The personalized speech recognition apparatus 102 may, in some embodiments, be embodied in various devices as described above. However, in some embodiments, the personalized speech recognition apparatus 102 may be embodied as a chip or chip set. In other words, the personalized speech recognition apparatus 102 may comprise one or more physical packages (e.g., chips) including materials, components and/or wires on a structural assembly (e.g., a baseboard). The structural assembly may provide physical strength, conservation of size, and/or limitation of electrical interaction for component circuitry included thereon. The personalized speech recognition apparatus 102 may therefore, in some cases, may be configured to implement an embodiment of the present invention on a single chip or as a single “system on a chip.” As such, in some cases, a chip or chipset may constitute means for performing one or more operations described herein for personalizing speech recognition in devices.

The processor 20 may be embodied in a number of different ways. For example, the processor 20 may be embodied as one or more of various hardware processing means such as a coprocessor, a microprocessor, a controller, a digital signal processor (DSP), a processing element with or without an accompanying DSP, or various other processing circuitry including integrated circuits such as, for example, an application specific integrated circuit (ASIC) an field programmable gate array (FPGA), a microcontroller unit (MCU), a hardware accelerator, a special-purpose computer chip, or the like. As such, in some embodiments, the processor 20 may include one or more processing cores configured to perform independently. A multi-core processor may enable multiprocessing within a single physical package. Additionally or alternatively, the processor 20 may include one or more processors configured in tandem via the bus to enable independent execution of instructions, pipelining and/or multithreading.

In an example embodiment, the processor 20 may be configured to execute instructions stored in the memory device 26 or otherwise accessible to the processor 20. In example embodiments, such instructions may provide for the retrieval, transmittal, and/or processing of SRMs, including generating additional SRMs based on received updated speaker dependent SRMs. Alternatively or additionally, the processor 20 may be configured to execute hard coded functionality. As such, whether configured by hardware or software methods, or by a combination thereof, the processor 20 may represent an entity (e.g., physically embodied in circuitry) capable of performing operations according to an embodiment of the present invention, such as the personalization of SRMs. Thus, for example, when the processor 20 is embodied as an ASIC, FPGA or the like, the processor 20 may be specifically configured hardware for conducting the operations described herein. Alternatively, as another example, when the processor 20 is embodied as an executor of software instructions, the instructions may specifically configure the processor 20 to perform the algorithms and/or operations described herein when the instructions are executed. However, in some cases, the processor 20 may be a processor of a specific device (e.g., a user terminal or network entity) configured to employ an embodiment of the present invention by further configuration of the processor 20 by instructions for performing the algorithms and/or operations described herein. The processor 20 may include, among other things, a clock, an arithmetic logic unit (ALU) and logic gates configured to support operation of the processor 20.

Meanwhile, the communication interface 24 may be any means such as a device or circuitry embodied in either hardware or a combination of hardware and software that is configured to receive and/or transmit data from/to a network and/or any other device or module in communication with the personalized speech recognition apparatus 102. In this regard, the communication interface 24 may include, for example, an antenna (or multiple antennas) and supporting hardware and/or software for enabling communications with a wireless communication network, for transmitting and receiving SRMs to and from remote devices. Additionally or alternatively, the communication interface 24 may include the circuitry for interacting with the antenna(s) to cause transmission of signals via the antenna(s) or to handle receipt of signals received via the antenna(s). In some environments, the communication interface 24 may alternatively or also support wired communication. As such, for example, the communication interface 24 may include a communication modem and/or other hardware/software for supporting communication via cable, digital subscriber line (DSL), universal serial bus (USB) or other mechanisms.

In some embodiments, such as instances in which the personalized speech recognition apparatus 102 is embodied by a user device, the personalized speech recognition apparatus 102 may include a user interface 22 that may, in turn, be in communication with the processor 20 to receive an indication of a user input and/or to cause provision of an audible, visual, mechanical or other output to the user. As such, the user interface 22 may include, for example, a keyboard, a mouse a display, a touch screen(s), touch areas, soft keys, a microphone, a speaker, or other input/output mechanisms. Alternatively or additionally, the processor 20 may comprise user interface circuitry configured to control at least some functions of one or more user interface elements such as, for example, a speaker, ringer, microphone, display, and/or the like. The processor 20 and/or user interface circuitry comprising the processor 20 may be configured to control one or more functions of one or more user interface elements through computer program instructions (e.g., software and/or firmware) stored on a memory accessible to the processor 20 (e.g., memory device 26, and/or the like).

In some example embodiments, processor 20 may be embodied as, include, or otherwise control a speech personalization administrator 28 for providing personalized speech recognition. As such, the speech personalization administrator 28 may be embodied as various means, such as circuitry, hardware, a computer program product comprising computer readable program instructions stored on a computer readable medium (for example, memory device 26) and executed by a processing device (for example, processor 20), or some combination thereof. Speech personalization administrator 28 may be capable of communication with one or more of the processor 20, memory device 26, user interface 22, and communication interface 24. As such, the speech personalization administrator 28 may be configured to generate additional

SRMs, adaptable by a variety of user terminals and that may be based on speaker dependent SRMs, as described above and in further detail hereinafter.

Any number of user terminal(s) 110, such as 110A and 110B, may connect to personalized speech recognition apparatus 102 via a network 100. User terminal 110 may be embodied as a mobile terminal, such as personal digital assistants (PDAs), pagers, mobile televisions, mobile telephones, gaming devices, laptop computers, tablet computers, cameras, camera phones, video recorders, audio/video players, radios, global positioning system (GPS) devices, navigation devices, or any combination of the aforementioned, and other types of devices capable of providing speech recognition. The user terminal 110 need not necessarily be embodied by a mobile device and, instead, may be embodied in a fixed device, such as a computer, workstation, or home appliance, such as a coffee maker. Additionally or alternatively, user terminal(s) 110 may be embodied in a vehicle, or any other machine or device capable of processing voice commands.

For simplicity, only user terminal 110A is illustrated in further detail, but it will be appreciated that any of the user terminals 110, such as user terminal 110B may be configured as illustrated in and described with respect to user terminal 110A. The user terminal 110 may therefore include or otherwise be in communication with processor 120, user interface 122, communication interface 124, and memory device 126.

In some embodiments, the processor 120 (and/or co-processors or any other processing circuitry assisting or otherwise associated with the processor 120) may be in communication with the memory device 126 via a bus for passing information among components of the user terminal 110. The memory device 126 may include, for example, one or more volatile and/or non-volatile memories. In other words, for example, the memory device 126 may be an electronic storage device (e.g., a computer readable storage medium) comprising gates configured to store data (e.g., bits) that may be retrievable by a machine (e.g., a computing device like the processor 120). The memory device 126 may be configured to store information, data, content, applications, instructions, or the like for enabling the user terminal to carry out various functions in accordance with an example embodiment of the present invention. For example, the memory device 126 could be configured to store SRMs, instructions for adapting SRMs with terminal dependent data, and instructions for training SRMs with speaker dependent data. Memory device 126 may therefore buffer input data for processing by the processor 120. Additionally or alternatively, the memory device 26 could be configured to store instructions for execution by the processor 120.

The processor 120 may be embodied in a number of different ways. For example, the processor 120 may be embodied as one or more of various hardware processing means such as a coprocessor, a microprocessor, a controller, a DSP, a processing element with or without an accompanying DSP, or various other processing circuitry including integrated circuits such as, for example, an ASIC , an FPGA, an MCU, a hardware accelerator, a special-purpose computer chip, or the like. As such, in some embodiments, the processor 120 may include one or more processing cores configured to perform independently. A multi-core processor may enable multiprocessing within a single physical package. Additionally or alternatively, the processor 120 may include one or more processors configured in tandem via the bus to enable independent execution of instructions, pipelining and/or multithreading.

In an example embodiment, the processor 120 may be configured to execute instructions stored in the memory device 126 or otherwise accessible to the processor 120. For example, the processor 120 may be configured to adapt an SRM advantageously to the user terminal, based on terminal dependent data, such as microphone information and context, so that the SRM may account for variances across user terminals. In example embodiments, the user terminal(s) 110 may include means, such as a processor 120, for training the SRM with speech input, to generate and/or refine a speaker dependent SRM that may improve speech input processing on the user terminal (and subsequently, other user terminals). Alternatively or additionally, the processor 120 may be configured to execute hard coded functionality. As such, whether configured by hardware or software methods, or by a combination thereof, the processor 120 may represent an entity (e.g., physically embodied in circuitry) capable of performing operations according to an embodiment of the present invention while configured accordingly. Thus, for example, when the processor 120 is embodied as an ASIC, FPGA or the like, the processor 120 may be specifically configured hardware for conducting the operations described herein. Alternatively, as another example, when the processor 120 is embodied as an executor of software instructions, the instructions may specifically configure the processor 120 to perform the algorithms and/or operations, such as adaptation and training of SRMs, processing of speech input, such as by using the SRMs, for conversion to text, when the instructions are executed. However, in some cases, the processor 120 may be a processor of a specific device (e.g., a mobile terminal or network entity) configured to employ an embodiment of the present invention by further configuration of the processor 120 by instructions for performing the algorithms and/or operations described herein. The processor 120 may include, among other things, a clock, an arithmetic logic unit (ALU) and logic gates configured to support operation of the processor 120.

Meanwhile, the communication interface 124 may be any means such as a device or circuitry embodied in either hardware or a combination of hardware and software that is configured to receive and/or transmit data from/to a network and/or any other device or module in communication with the user terminal 110. In example embodiments, the communication interface 124 may be specifically configured for transmitting and receiving SRMs to and from the personalized speech recognition apparatus 102. In this regard, the communication interface

124 may include, for example, an antenna (or multiple antennas) and supporting hardware and/or software for enabling communications with a wireless communication network. Additionally or alternatively, the communication interface 124 may include the circuitry for interacting with the antenna(s) to cause transmission of signals via the antenna(s) or to handle receipt of signals received via the antenna(s). In some environments, the communication interface 124 may alternatively or also support wired communication for communication of SRMs. As such, for example, the communication interface 124 may include a communication modem and/or other hardware/software for supporting communication via cable, digital subscriber line (DSL), universal serial bus (USB) or other mechanisms.

The user terminal 110 may include a user interface 122 that may, in turn, be in communication with the processor 120 to receive an indication of a user input and/or to cause provision of an audible, visual, mechanical or other output to the user. As such, the user interface 122 may include, for example, a keyboard, a mouse, a display, a touch screen(s), touch areas, soft keys, a microphone, a speaker, or other input/output mechanisms. The user interface 122 may therefore be configured to receive speech input, such as, via a microphone, for the purposes of speech recognition and/or training of an SRM. Alternatively or additionally, the processor 120 may comprise user interface circuitry configured to control at least some functions of one or more user interface elements such as, for example, a speaker, ringer, microphone, display, and/or the like. The processor 120 and/or user interface circuitry comprising the processor 120 may be configured to control one or more functions of one or more user interface elements through computer program instructions (e.g., software and/or firmware) stored on a memory accessible to the processor 120 (e.g., memory device 126, and/or the like).

Network 100 may be embodied in a local area network, the Internet, any other form of a network, or in any combination thereof, including proprietary private and semi-private networks and public networks. The network 100 may comprise a wire line network, wireless network (e.g., a cellular network, wireless local area network, wireless wide area network, some combination thereof, or the like), or a combination thereof, and in some example embodiments comprises at least a portion of the Internet. The network 100 may be used for transmitting speaker dependent data and/or SRMs to and from devices. As another example, a user terminal 110 may be directly coupled to and/or may include a personalized speech recognition apparatus 102.

Referring now to Figure 2, the operations for receiving and adapting an SRM on a user terminal, in accordance with one embodiment of the present invention are outlined in accordance with one example embodiment. In this regard and as described below, the operations of Figures 2 may be performed by the user terminal 110A, user terminal 110B, and/or the like, for example.

As shown by operation 200, the user terminal 110A may include means, such as the processor 120, communication interface 124, or the like, for receiving at least one portion of an SRM, wherein the at least one portion of an SRM is stored remotely and is adaptable by one of more user terminals to process input speech. In other words, the user terminal 110A may receive at least one portion of an SRM from the personalized speech recognition apparatus 102, for example, including any combination of the HMM, DTW, NN, and FST models, as described above. The at least one portion of an SRM may also include any combination of speaker independent data and/or speaker dependent data, and may be adaptable by the user terminal 110A to process speech input (e.g., perform speech recognition tasks). The adaptation is described in further detail with respect to operation 210.

To receive the at least one portion of an SRM on the user terminal 110A, in an example embodiment, a user of user terminal 110A may provide logon credentials or the like, via user interface 122, communication interface 124, and/or network 100 to the personalized speech recognition apparatus 102. In some embodiments, the user terminal 110A may check for updates by communicating with the personalized speech recognition apparatus 102, and receive an SRM or portion thereof if an update is available. In some examples, an update may be available if a user updated, based on training, verification or the like on another device, such as user terminal 110B.

In some embodiments, the user terminal 110A may download an SRM or portion thereof for the first time (such as during initial device setup, or factory reset), or the newly received SRM or portion thereof may include updates compared to a previous version used by user terminal 110A. In some embodiments, receipt of the SRM or portion thereof by the user terminal 110A may occur during scheduled update routines that may be unobtrusive to or unnoticed by a user. That is, the synchronization may occur seamlessly as a background system update. Additionally or alternatively, a request for an SRM or portion thereof may be explicitly initiated on the user terminal 110A (such as logging onto the personalized speech recognition apparatus 102 and requesting an update). In some embodiments, an update may be initiated by the personalized speech recognition apparatus 102. For example, a user may be automatically notified that an update is available, such as by Short Message Service (SMS), for example, so as to confirm that they would like to receive the at least one portion of an SRM on the user terminal 110A.

The user terminal 110A may therefore receive at least one portion of an SRM associated with the individual user (such as identified with the logon credentials). Additionally or alternatively, the SRM or portion thereof may be identified by the personalized speech recognition apparatus by other means. For example, a user of a device may provide a geographic location, via a Global Positioning Device (GPS) and/or manual indication of a location, for example. The user terminal 110A may therefore receive an SRM based on a geographic location and /or dialect.

Having received at least one portion of an SRM, as described with respect to operation 200, the user terminal 110A may include means, such as the processor 120, for accessing a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of speech recognition model. As such, the received at least one portion of an SRM may be a complete SRM, and may therefore be stored on memory device 126, and accessed by the processor 120. In some embodiments, where the at least one portion of the SRM does not provide a complete or fully functioning SRM, the processor 120 may incorporate the at least one portion of an SRM to form a complete SRM. As such the SRM may be stored and accessed on memory device 126, for example.

Having accessed an SRM, as shown by operation 208, the user terminal 110A may include means, such as the processor 120, for adapting the SRM based on or more terminal dependent data. The terminal dependent data may include information regarding the user terminal 110A itself, such as characteristics of the microphone on the user terminal 110A to capture the speech input, and/or a context of the user terminal (e.g., an environment the device is commonly used in, or the intended purpose of the user terminal), or any settings of the user terminal 110A that could impact the process of speech input. The processor 120 may therefore utilize the terminal dependent data in adapting the SRM for use on the user terminal 110A.

In an example embodiment, microphone information may be retrieved from memory device 126, or read from a microphone component of user interface 122 by processor 120, for example. The microphone information may include any information relating to the microphone that may impact how speech input is recognized and/or processed according to the SRM. For example, the microphone information may comprise a microphone model identifier, or orientation of the microphone within the device. The microphone may additionally or alternatively be characterized by its transduction type, such as condenser and/or dynamic, for example. The user terminal 110A, using the processor 120, may therefore adapt the SRM according to microphone information to account for acoustic, phonetic, and/or other variances between microphones. For example, calculations in a DTW model may be consistently modified throughout, so that the user terminal 110A may accurately interpret sounds captured by the microphone.

In another example embodiment, the user terminal 110A may adapt the SRM based on the context of the user terminal. Use of an SRM by a speaker phone in a vehicle, for example, may be subject to background noise, such as wind, and/or radio or other device interference. The processor 120 of user terminal 110A may therefore adapt the received SRM, which in its previous state may not have accounted for such background noises, accordingly. Information regarding the context or use of the user terminal 110A may be explicitly retrieved from memory device 126, for example, and/or derived from various components of the user terminal 110A, allowing processor 120 to adapt the SRM based on what contexts the user terminal 110A will most likely be used in.

Although microphone information and context of the user terminal are provided as example terminal dependent data, it will be appreciated that numerous other terminal dependent data exist. Settings configuring various components of the user terminal 110A may be considered by the processor 120 in adapting the SRM for the user terminal 110A. In some embodiments, the settings may affect the adaptation of the SRM, and/or cause the processor 120 to adjust the settings of the user terminal 110A to tailor the device for use of the SRM. An adapted SRM may be stored on memory device 126, for example.

As shown by operation 220, the user terminal 110A may include means, such as the user interface 122, communication interface 124, and/or processor 120 for receiving a speech input. The speech input may be provided by a user to user terminal 110B by using a microphone of user interface 122, for example.

Additionally or alternatively, the user terminal 110A may receive a speech input through everyday use of the user terminal and may process the speech to generate text. The user terminal 110A may process received speech input using the SRM, and generate a textual output. In some examples, the processor 120 may process the speech input according to the SRM. For example, the processor 120 may calculate observation probability on the speech input based on the SRM that includes one or more HMM, DTW, NN, or FST models, for example. By way of further example, the processor 120 may identify a reference word with the highest probability when compared to other reference words, a threshold or the like. Based on those probabilities, the processor may then select or otherwise generate the speech recognition result (e.g. a text output).

As shown by operation 230, the user terminal 110A may include means, such as the user interface 122, communication interface 124 and/or processor 120, for verifying or correcting a processing of the speech input. The verification or correction could be received explicitly by a user input to the user terminal 110A, or implicitly by everyday use of the user terminal 110A.

For example, the user terminal 110A may be configured to receive an explicit correction of a processed speech input. In applications employing speech recognition, such as an example application that prefills dictated words in a draft email message, the interpretation of the speech input may be incorrect. In such cases, the user may correct a misinterpreted word(s) by selecting the misinterpreted word, and typing the corrected word in its place. See Figure 4.

As is provided in Figure 4, a user interface 122 may display an indication 400 of a word, such as a word that is misinterpreted, such as a word that is misinterpreted during the processing of input speech. In some examples, indication 400 may be provided by the user terminal 110A in scenarios such as those in which the SRM provided no reference word above some threshold probability, indicating that the processing of the speech input was not likely correct. Additionally or alternatively, the indication 400 may be provided explicitly by a user, by selection of the word for correction, for example. User input 410 provides a means for receiving a correction of the processed speech input. In this example, the speech recognition system has interpreted the word "forest," and a user provides the correct phrase, "for the rest."

In other examples a speech input may be deemed as correct based on implicit verification. For example, a user terminal, such as user terminal 110A, may be embodied as a mobile phone and may further be operable to receive a speech input such as "call Suzanne." Upon automatic selection and execution of the associated command (e.g., initiating a call to a phone number saved for a contact by the name of Suzanne), and failure to receive any correction to stop the initiated phone call, the user terminal 100A, such as by the processor 120, may consider this absence of any action by the user a verification of the processed speech input.

As shown by operation 240, the user terminal 110A, such as by processor 120, and memory device 126, for example, may generate and/or otherwise refine a speaker dependent SRM based on the speech input. As such, the SRM may be trained using speech input received with respect to operation 220, and/or verification or correction of the processed speech input with respect to operation 230. Existing SRMs on memory device 126 may therefore be tailored for use by a particular user or group of users. Additionally or alternatively, new speaker dependent SRMs may be generated for improved speech input processing.

Training can be performed, for example, by using feature vectors of the speech input (provided with respect to operation 220) and associating them with corresponding reference words, as provided by the verification and/or correction with respect to the operation 230 above. Additionally or alternatively, a verification or correction need not be provided, but the processor 120 may identify the reference words from a script on memory device 126 (such as in an example embodiment where the speech input is received based on a script).

The SRM, such as an HMM, DTW, NN, FST, or the like, may therefore be expanded, or otherwise modified, to incorporate the speech input and associated reference words. In some examples, processed speech input and associated reference words may be further processed by processor 120, and applied to an existing SRM, to refine a speaker dependent SRM. In some embodiments, where an SRM is not already present on the user terminal 110A, a new speaker dependent SRM may be generated. The generated or refined speaker dependent SRM may be stored on memory device 126, for example.

As shown by operation 250, the user terminal 110A may include means, such as communication interface 124, and/or processor 120, for causing transmission of the speaker dependent SRM to a remote storage location, such as personalized speech recognition apparatus 102, for example. Transmission of the speaker dependent SRM to a remote location may allow the speaker dependent SRM to be advantageously transmitted to other user terminals, such as described in further detail with respect to Figure 3. Further, and in some examples, by transmitting the speaker dependent SRM to the remote location, one or more user terminals may provide updates to or otherwise refine the speaker dependent SRM. The speaker dependent SRM may therefore be retrieved from memory device 126, and transmitted via communication interface 124 and over network 100, for example, to the remote storage location.

In some embodiments, the transmission may occur automatically following generation and/or refinement of the speaker dependent SRM with respect to operation 240. In some embodiments, a user of user terminal 110A may initiate the transmission, such as for example, providing logon credentials to the personalized speech recognition apparatus 102, as described with respect to operation 200, for example. The speaker dependent SRM may then be transmitted to the personalized speech recognition apparatus 102 for storage, and subsequent retrievals.

Figure 3 is a flowchart illustrating operations to transmit an SRM, receive a speaker dependent SRM, and generate an additional SRM, using a speech personalization apparatus 102 in accordance with one embodiment of the present invention.

As shown by operation 300, the personalized speech recognition apparatus 102 may include means, such as the processor 20, speech personalization administrator 28, communication interface 24, or the like, for causing transmission of an SRM (or portion thereof) to a user terminal. The SRM may therefore be retrieved from memory device 26, and sent over network 110, via communication interface 24, to user terminal 110A, for example. In some examples, the SRM that is transmitted may be an SRM that is configured for a particular device, a particular region or dialect or the like.

For example, the personalized speech recognition apparatus 102 may generate the additional SRM based on an associated with a group of users, such as one associated with a geographic location. For example, some geographic areas, like the southern United States, for example, may experience regional accents that may otherwise confuse speech input processing systems. Personalized speech recognition apparatus 102 may therefore generate the additional SRM based on a particular geographic location in order to subsequently provide more accurate speech recognition functions to users in, from, or otherwise associated with the same geographic location.

Similarly, an additional SRM may be generated based on a specific dialect. For example, due to varying dialects, some words may be pronounced differently than the same word in a different language, potentially causing erroneous speech input processing on a user terminal. Personalized speech recognition apparatus 102 may therefore associate the speaker dependent SRM with a dialect in order to provide more accurate speech recognition functions to users whose speech is closely related to the specific dialect. A user of a device may then provide indication of a particular dialect, and receive an SRM adapted for that dialect.

Alternatively or additionally, the SRM may already be adapted to a particular user. For example, the personalized speech recognition apparatus 102 may receive logon information from a user terminal, such as user terminal 110A that indicates the identity of a particular user. As such, personalized speech recognition apparatus 102, such as via the processor 20, the communications interface 24 or the like, may cause the SRM related to the particular user to be transmitted to user terminal 110A.

The transmission may be initiated on the personalized speech recognition apparatus 102 in various ways, such as receiving requests initiated explicitly (e.g., logon) or automatically (e.g., initial installation) from the user terminal 110A, and/or automatic transmission imitated by the personalized speech recognition apparatus 102. Various other methods for initiation of transmission of the SRM are described herein.

As shown by operation 310, the personalized speech recognition apparatus 102 may include means, such as the processor 20, speech personalization administrator 28, communication interface 24, or the like, for receiving at least a portion of a speaker dependent SRM from the user terminal, such as user terminal 110A. In some examples, the received speaker dependent SRM (or portion thereof) may contain one or more updates to or refinements of the speaker dependent SRM as is described with respect to operations 240 and 250 of Figure 2.

As shown by operation 320, the personalized speech recognition apparatus 102 may include means, such as the processor 20, speech personalization administrator 28, or the like, for generating an additional or otherwise updated SRM based on the speaker dependent SRM, wherein the additional SRM is adaptable by one or more user terminals. In some examples, the additional SRM is contrasted based on the speaker dependent SRM and comprises the updates to or refinements of the SRM from the user terminal, as well as, one or more other user terminals.

As such, the speech personalization administrator 28 may access an existing SRM on memory 26, and modify, update or otherwise refine the SRM with the speaker dependent SRM, or a portion of the speaker dependent SRM, accordingly. Additionally, or alternatively, a new SRM may be generated using the speaker dependent SRM. The additional SRM may be, or otherwise include, a HMM, DTW, NN, or FST, for example. The additional SRM may be adaptable by one or more user terminals, such as described with respect to operations 200 and 210 above.

As shown by operation 330, the personalized speech recognition apparatus 102 may include means, such as the processor 20, communication interface 24, or the like, for causing transmission of the additional SRM to an additional device. The transmission may be initiated and completed by use of similar operations described with respect to operation 300, but the SRM may this time be transmitted to a different terminal, such as user terminal 110B, for example. Advantageously, for example, the additional SRM may be shared between one or more user terminals, devices and/or the like.

In an example embodiment, the personalized speech recognition apparatus 102 may select the additional SRM to transmit to the user terminal 110B, based on a variety of factors, such as terminal dependent data, and/or user identification, for example. An association of the individual user (or group of users) and speaker dependent SRM may allow the personalized speech recognition apparatus 102 to advantageously provide the SRM on demand, to various devices belonging to a user.

For example, a user terminal 110A embodied as a personal computer or laptop capable of producing text from speech input, such as dictated reports or emails, may rely on an extensive SRM representing an entire natural language. A speaker dependent SRM generated and/or refined on the user terminal 110A may be available on personalized speech recognition apparatus 102 for distribution to one or more other user terminals.

For example, if the same user of user terminal 110A purchases a new user terminal 110B, it may be advantageous to provide portions of the speaker dependent SRM to the user terminal 110B. Presume for example that the user terminal 110B is a coffee maker. A coffee maker, may not require such a broad vocabulary required by the personal computer or laptop embodiment of user terminal 110A, but only portions of the speaker dependent SRM including language relating to functions of the coffee maker (e.g., grind, brew), or language relating to measurements and/or timing. As such, upon detecting that user terminal 110B is embodied as a coffee maker, for example, the personalized speech recognition apparatus 102 may advantageously select an SRM or a portion of an SRM (as generated in operation 320) for use by the coffee maker, potentially minimizing bandwidth required for transmitting, and memory required for storing (on user terminal 110B), an otherwise extensive SRM.

Having received an SRM from the personalized speech recognition apparatus 102, the user terminal 110B may utilize the SRM in processing of speech input, therefore offering to its users personalized speech recognition or in other examples, by not needing its users to retrain a SRM.

That is, the speech input processing may be improved by use of the SRM, which may include a portion(s) of the speaker dependent SRM generated or refined on user terminal 110A. As such, the user of user terminal 110B may provide speech input to user terminal 110B, and experience reduced or minimized error rates in speech input processing and/or execution of associated voice commands.

It will be appreciated that, although Figure 1 illustrates an embodiment utilizing user terminals 110A and 110B, and a personalized speech recognition apparatus 102, many other configurations exist. Indeed, a personalized speech recognition apparatus 102 may be locally installed on a device such as user terminal 110A and/or 110B and configured to run independently, where data may not necessarily be shared across devices or a server.

In some embodiments, a user terminal such as user terminal 110A and/or 110B may provide a speaker dependent SRM to a personalized speech recognition apparatus 102, may receive an SRM from a personalized speech recognition apparatus 102, or may both provide and receive the same, respectively. In such example embodiments, the personalized speech recognition apparatus 102 may be implemented in the cloud, and data may be transmitted between user terminal(s) and server(s) over network 100. In a particularly advantageous embodiment, various user terminals may routinely receive updated SRMs, thereby continually improving speech input processing by utilizing speaker dependent SRMs generated and/or refined on other user terminals.

Additionally or alternatively, in some embodiments, a device such as user terminal 110A and/or 110B may be shipped with SRMs preinstalled. In some embodiments, the SRM may be local to the area or country the user terminal is distributed in. That is, the SRM may be based on a speaker dependent SRM based on a dialect or geographic area.

As described above, Figures 2 and 3 are flowcharts illustrating operations performed by a user terminal 110A user terminal 110B and/or the like, and personalized speech recognition apparatus 102, respectively. It will be understood that each block of the flowchart, and combinations of blocks in the flowcharts, may be implemented by various means, such as hardware, firmware, processor, circuitry, and/or other devices associated with execution of software including one or more computer program instructions. For example, one or more of the procedures described above may be embodied by computer program instructions. In this regard, the computer program instructions which embody the procedures described above may be stored by a memory device 26 or 126 employing an embodiment of the present invention and executed by a processor 20 or 120. As will be appreciated, any such computer program instructions may be loaded onto a computer or other programmable apparatus (e.g., hardware) to produce a machine, such that the resulting computer or other programmable apparatus implements the functions specified in the flowcharts' blocks. These computer program instructions may also be stored in a computer-readable memory that may direct a computer or other programmable apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture the execution of which implements the function specified in the flowcharts' blocks. The computer program instructions may also be loaded onto a computer or other programmable apparatus to cause a series of operations to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions which execute on the computer or other programmable apparatus provide operations for implementing the functions specified in the flowcharts' blocks.

Accordingly, blocks of the flowcharts support combinations of means for performing the specified functions and combinations of operations for performing the specified functions for performing the specified functions. It will also be understood that one or more blocks of the flowcharts, and combinations of blocks in the flowcharts, may be implemented by special purpose hardware-based computer systems which perform the specified functions, or combinations of special purpose hardware and computer instructions.

In some embodiments, certain ones of the operations above may be modified or further amplified. Furthermore, in some embodiments, additional optional operations may be included as indicated by the blocks shown with a dashed outline in Figures 2 and 3. Modifications, additions, or amplifications to the operations above may be performed in any order and in any combination.

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Moreover, although the foregoing descriptions and the associated drawings describe example embodiments in the context of certain example combinations of elements and/or functions, it should be appreciated that different combinations of elements and/or functions may be provided by alternative embodiments without departing from the scope of the appended claims. In this regard, for example, different combinations of elements and/or functions than those explicitly described above are also contemplated as may be set forth in some of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

THAT WHICH IS CLAIMED

1. A method comprising:

   receiving at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely and is adaptable by one or more user terminal to process input speech;

   accessing a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model; and

   adapting the speech recognition model based on terminal dependent data.

2. A method according to claim 1, further comprising:

   processing received speech input using the speech recognition model; and

   generating a textual output.

3. A method according to claim 1 or 2, further comprising:

   receiving a speech input; and

   refining a speaker dependent speech recognition model based on the speech input.

4. A method according to claim 3, further comprising:

   verifying or correcting a processing of the speech input, wherein refining the speaker dependent speech recognition model is further based on the verification or correction.

5. A method according to claim 3, or 4, further comprising:

   causing transmission of at least one portion of the speaker dependent speech recognition model to a remote storage location.

6. The method according to claim 1, 2, 3, or 4, wherein the terminal dependent data comprises microphone information.

7. The method according to claim 1, 2, 3, 4 or 5, wherein the terminal dependent data comprises a context.

8. The method according to claim 1, 2, 3, 4, 5, or 6, wherein the received at least one portion of a speech recognition model is received based on one of at least an individual user, group of users, geographic location, or a dialect.

9. A method according to claim 1, 2, 3, 4, 5, 6 or 7, wherein the received at least one portion of a speech recognition model is based on at least one of a Hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

10. A method comprising:

receiving at least one portion of a speaker dependent speech recognition model from a user terminal; and

generating at least one additional portion of a speech recognition model based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of a speech recognition model is adaptable by one or more user terminals.

11. A method according to claim 10, further comprising:

causing transmission of the at least one additional portion of the speech recognition model to an additional user terminal.

12. A method according to claim 10 or 11, wherein generating the at least one additional portion of the speech recognition model is further based on one of at least an individual user, group of users, geographic location, or a dialect.

13. A method according to claim 10, 11 or 12, wherein the at least one additional portion of the speech recognition model is based on at least one of a hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

14. An apparatus comprising at least one processor and at least one memory including computer program code, the at least one memory and the computer program code configured to, with the processor, cause the apparatus to at least:

receive at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely, and is adaptable by one or more user terminals to process input speech;

access a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model; and

adapt the speech recognition model based on terminal dependent data.

15. An apparatus according to claim 14, wherein the at least one memory and the computer program code are further configured to, with the processor, cause the apparatus to at least:

process received speech input using the speech recognition model; and

generate a textual output.

16. An apparatus according to claim 14 or 15, wherein the at least one memory and the computer program code are further configured to, with the processor, cause the apparatus to at least:

    receive a speech input; and

    refine a speaker dependent speech recognition model based on the speech input.

17. An apparatus according to claim 16, wherein the at least one memory and the computer program code are further configured to, with the processor, cause the apparatus to at least:

    verify or correct a processing of the speech input, wherein refining the speaker dependent speech recognition model is further based on the verification or correction.

18. An apparatus according to claim 16 or 17, wherein the at least one memory and the computer program code are further configured to, with the processor, cause the apparatus to at least:

    cause transmission of at least a portion of the speaker dependent speech recognition model to a remote storage location.

19. An apparatus according to claim 14, 15, 16, 17 or 18 wherein the terminal dependent data comprises microphone information.

20. An apparatus according to claim 14, 15, 16, 17, 18 or 19, wherein the terminal dependent data comprises a context.

21. An apparatus according to claim 14, 15, 16, 17 or 18, wherein the received at least one portion of a speech recognition model is received based on one of at least an individual user, group of users, geographic location, or a dialect.

22. An apparatus according to claim 14, 15, 16, 17, 18, 19, 20 or 21, wherein the received at least one portion of a speech recognition model is based on at least one of a Hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

23. An apparatus comprising at least one processor and at least one memory including computer program code, the at least one memory and the computer program code configured to, with the processor, cause the apparatus to at least:

    receive at least one portion of a speaker dependent speech recognition model from a user terminal; and

    generate at least one additional portion of a speech recognition model based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of the speech recognition model is adaptable by one or more user terminals.

24. An apparatus according to claim 23, wherein the at least one memory and the computer program code are further configured to, with the processor, cause the apparatus to at least:

    cause transmission of the at least one additional portion of a speech recognition model to an additional user terminal.

25. An apparatus according to claim 23 or 24, wherein generating the at least one additional portion of a speech recognition model is further based on one of at least an individual user, group of users, geographic location, or a dialect.

26. An apparatus according to claim 23, 24 or 25, wherein the at least one additional portion of the speech recognition model is based on at least one of a hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

27. A computer program product comprising at least one non-transitory computer-readable storage medium having computer-executable program code instructions stored therein, the computer-executable program code instructions comprising program code instructions to:

    receive at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely, and is adaptable by one or more user terminals to process input speech;

    access a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model; and

    adapt the speech recognition model based on terminal dependent data.

28. A computer program product according to claim 27, wherein the computer-executable program code instructions further comprise program code instructions to:

    process received speech input using the speech recognition model; and

    generate a textual output.

29. A computer program product according to claim 27 or 28, wherein the computer-executable program code instructions further comprise program code instructions to:

    receive a speech input; and

    refine a speaker dependent speech recognition model based on the speech input.

30. A computer program product according to claim 29, wherein the computer-executable program code instructions further comprise program code instructions to:

    verify or correct a processing of the speech input, wherein refining the speaker dependent speech recognition model is further based on the verification or correction.´

31. A computer program product according to claim 29 or 30, wherein the computer-executable program code instructions further comprise program code instructions to:

    cause transmission of at least one portion of the speaker dependent speech recognition model to a remote storage location.

32. A computer program product according to claim 27, 28, 29, 30 or 31, wherein the terminal dependent data comprises microphone information.

33. A computer program product according to claim 27, 28, 29, 30, 31 or 32, wherein the terminal dependent data comprises a context.

34. A computer program product according to claim 27, 28, 29, 30, 31, 32 or 33, wherein the received at least one portion of a speech recognition model is received based on one of at least an individual user, group of users, geographic location, or a dialect.

35. A computer program product according to claim 27, 28, 29, 30, 31, 32, 33 or 34, wherein the received at least one portion of a speech recognition model is based on at least one of a Hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

36. A computer program product comprising at least one non-transitory computer-readable storage medium having computer-executable program code instructions stored therein, the computer-executable program code instructions comprising program code instructions to:

    receive at least one portion of a speaker dependent speech recognition model from a user terminal; and

    generate at least one additional portion of a speech recognition model based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of a speech recognition model is adaptable by one or more user terminals.

37. A computer program product according to claim 36, wherein the computer-executable program code instructions further comprise program code instructions to:

cause transmission of the at least one additional portion or the additional speech recognition model to an additional user terminal.

38. A computer program product according to claim 36 or 37, wherein generating the at least one additional portion of the speech recognition model is further based on one of at least an individual user, group of users, geographic location, or a dialect.

39. A computer program product according to claim 36, 37 or 38, wherein the at least one additional portion of the speech recognition model is based on at least one of a hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

40. An apparatus comprising means for:

receiving at least one portion of a speech recognition model, wherein the at least one portion of the speech recognition model is stored remotely, and is adaptable by one or more user terminals to process input speech;

accessing a speech recognition model, wherein the speech recognition model is based on at least the received at least one portion of the speech recognition model; and

adapting the speech recognition model based on terminal dependent data.

41. An apparatus according to claim 40, further comprising means for:

process received speech input using the speech recognition model; and

generate a textual output.

42. An apparatus according to claim 40 or 41, further comprising means for:

receiving a speech input; and

refining a speaker dependent speech recognition model based on the speech input.

43. An apparatus according to claim 42, further comprising means for:

verifying or correcting a processing of the speech input, wherein refining the speaker dependent speech recognition model is further based on the verification or correction.

44. An apparatus according to claim 42 or 43, further comprising means for:

causing transmission of at least one portion of the speaker dependent speech recognition model to a remote storage location.

45. An apparatus according to claim 40, 41, 42, 43 or 44 wherein the terminal dependent data comprises microphone information.

46. An apparatus according to claim 40, 41, 42, 43, 44 or 45, wherein the terminal dependent data comprises a context.

47. An apparatus according to claim 40, 41, 42, 43, 44, 45 or 46, wherein the received at least one portion of a speech recognition model is received based on one of at least an individual user, group of users, geographic location, or a dialect.

48. An apparatus according to claim 40, 41, 42, 43, 44, 45, 46 or 47, wherein the received at least one portion of a speech recognition model is based on at least one of a Hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

49. An apparatus comprising means for:

receiving at least one portion of a speaker dependent speech recognition model from a user terminal; and

generating at least one additional portion of a speech recognition based on the received at least one portion of a speaker dependent speech recognition model, wherein the at least one additional portion of the speech recognition model is adaptable by one or more user terminals.

50. An apparatus according to claim 49, further comprising means for:

causing transmission of the at least one additional portion to an additional user terminal.

51. An apparatus according to claim 49 or 50, wherein generating the at least one additional portion of the speech recognition model is further based on one of at least an individual user, group of users, geographic location, or a dialect.

52. An apparatus according to claim 49, 50 or 51, wherein the at least one additional portion or the additional speech recognition model is based on at least one of a hidden Markov Model, dynamic time warping model, neural network, or finite state transducer.

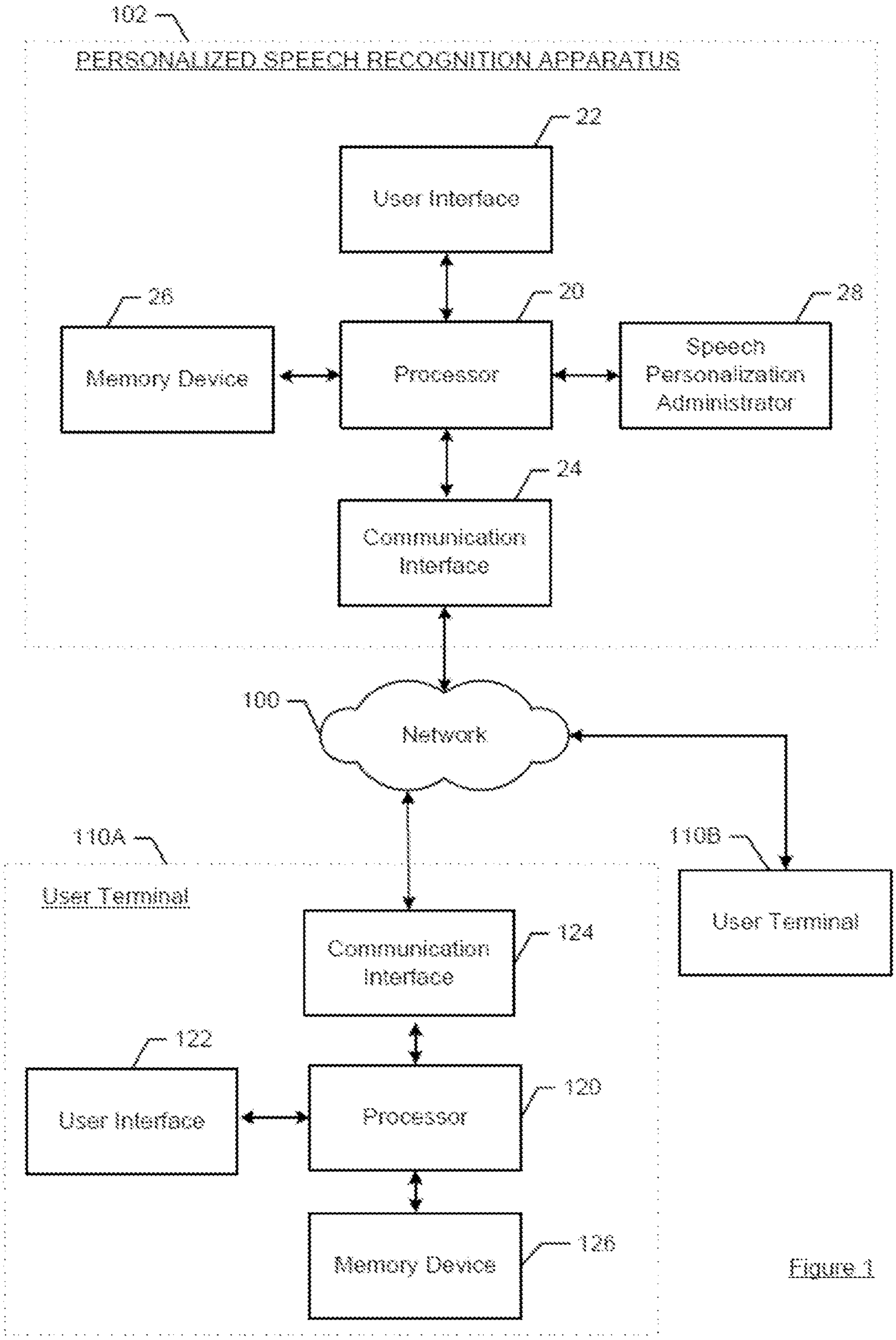
102
PERSONALIZED SPEECH RECOGNITION APPARATUS
22
User Interface
26
Memory Device
20
Processor
28
Speech Personalization Administrator
24
Communication Interface
100
Network
110A
User Terminal
124
Communication Interface
122
User Interface
120
Processor
126
Memory Device
110B
User Terminal

Figure 1

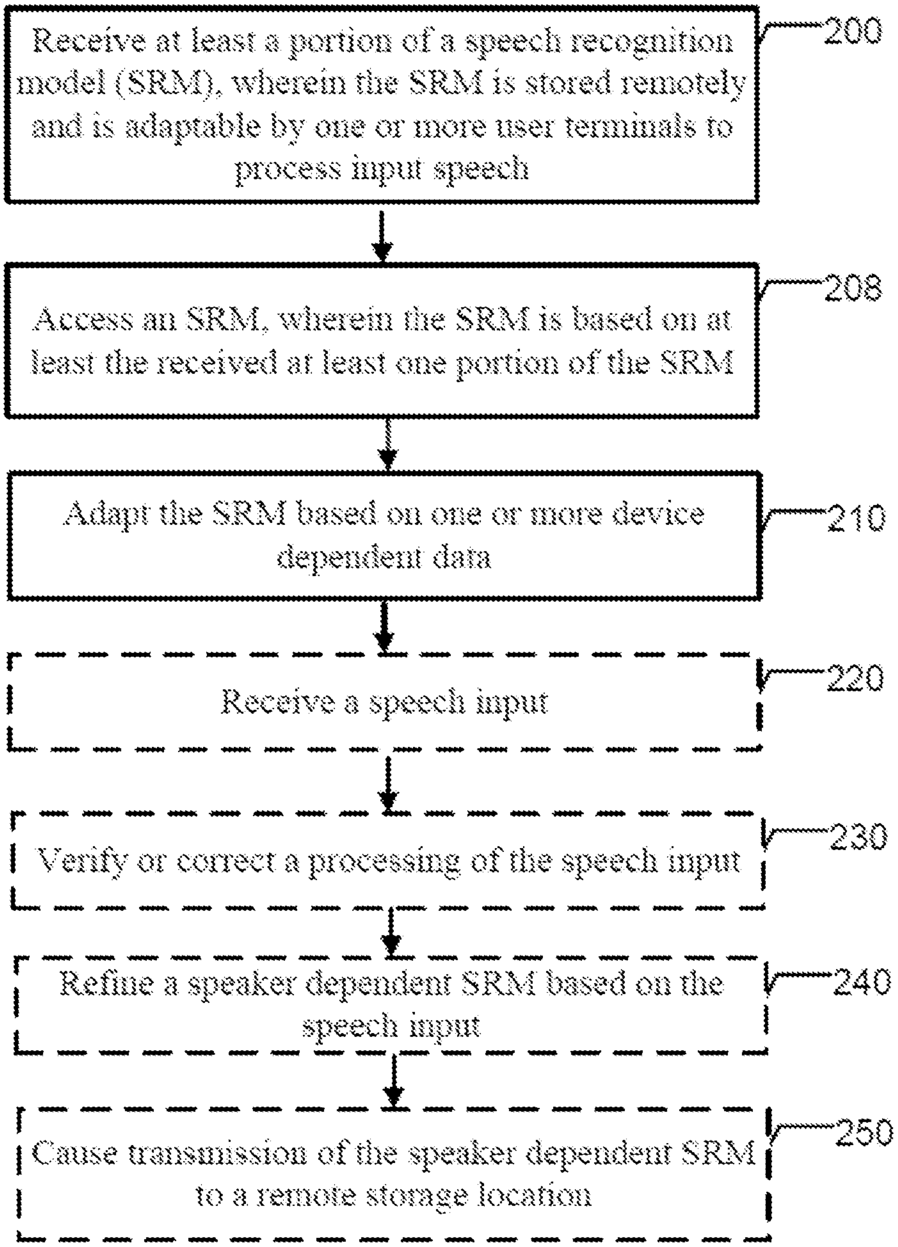
Receive at least a portion of a speech recognition model (SRM), wherein the SRM is stored remotely and is adaptable by one or more user terminals to process input speech
200
Access an SRM, wherein the SRM is based on at least the received at least one portion of the SRM
208
Adapt the SRM based on one or more device dependent data
210
Receive a speech input
220
Verify or correct a processing of the speech input
230
Refine a speaker dependent SRM based on the speech input
240
Cause transmission of the speaker dependent SRM to a remote storage location
250

Figure 2

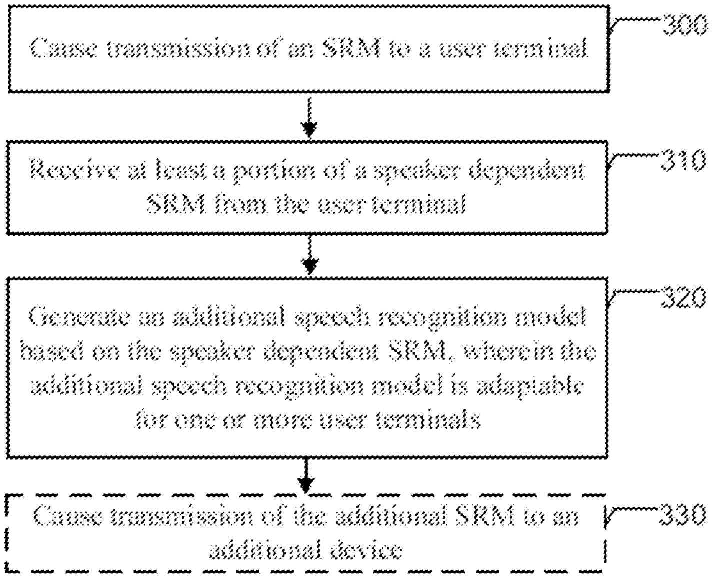
Cause transmission of an SRM to a user terminal
300
Receive at least a portion of a speaker dependent SRM from the user terminal
310
Generate an additional speech recognition model based on the speaker dependent SRM, wherein the additional speech recognition model is adaptable for one or more user terminals
320
Cause transmission of the additional SRM to an additional device
330

Figure 3

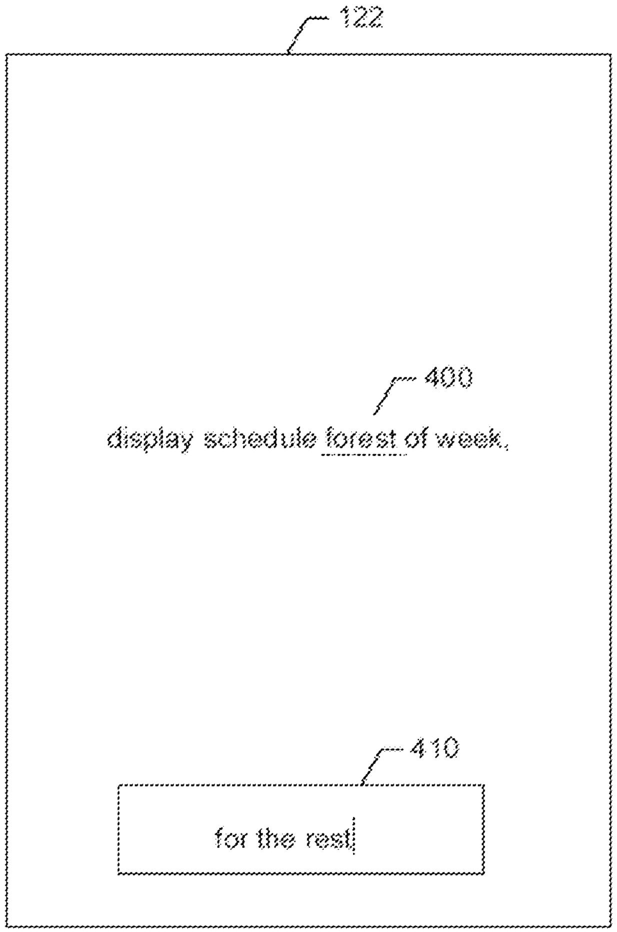

Figure 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2012/051285

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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)

IPC: G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

FI, SE, NO, DK

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI, XPESP, XPESP2, XPI3E, XPRD, INSPEC, NPL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2007124134 A1 (VAN KOMMER ROBERT [CH]) 31 May 2007 (31.05.2007) Fig. 1; paragraphs [0033]-[0046]	1-52
X	US 2003050783 A1 (YOSHIZAWA SHINICHI [JP]) 13 March 2003 (13.03.2003) Fig. 10; paragraphs [0156]-[0158], [0163], [0165]-[0169], [0177], [0181]	1-9, 14-22, 27-35, 40-48
A	US 2010145699 A1 (TIAN JILEI [FI]) 10 June 2010 (10.06.2010) abstract; Figs. 1, 2	

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* 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 but 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 15 October 2013 (15.10.2013)	Date of mailing of the international search report 18 October 2013 (18.10.2013)
Name and mailing address of the ISA/FI National Board of Patents and Registration of Finland P.O. Box 1160, FI-00101 HELSINKI, Finland Facsimile No. +358 9 6939 5328	Authorized officer Vesa-Matti Louekoski Telephone No. +358 9 6939 500

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/FI2012/051285

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
US 2007124134 A1	31/05/2007	AT 439665 T	15/08/2009
		DE 602005015984 D1	24/09/2009
		EP 1791114 A1	30/05/2007
		EP 1791114 B1	12/08/2009
		EP 2109097 A1	14/10/2009
		ES 2330758 T3	15/12/2009
		US 8005680 B2	23/08/2011
US 2003050783 A1	13/03/2003	CN 1409527 A	09/04/2003
		EP 1293964 A2	19/03/2003
		JP 2003177790 A	27/06/2003
		JP 2005107550 A	21/04/2005
US 2010145699 A1	10/06/2010	CA 2745991 A1	17/06/2010
		CN 102282608 A	14/12/2011
		CN 102282608 B	12/06/2013
		EP 2356651 A1	17/08/2011
		JP 2012511730 A	24/05/2012
		JP 5172021 B2	27/03/2013
		KR 20110100642 A	14/09/2011
		US 8155961 B2	10/04/2012
		WO 2010067165 A1	17/06/2010

CLASSIFICATION OF SUBJECT MATTER

Int.Cl.
G10L 15/065 (2013.01)
G10L 15/30 (2013.01)