



- (51) **International Patent Classification:**
G10L 15/14 (2006.01) *G10L 15/183* (2013.01)
G10L 15/30 (2013.01)
- (21) **International Application Number:**
 PCT/US2016/046162
- (22) **International Filing Date:**
 9 August 2016 (09.08.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
 14/864,456 24 September 2015 (24.09.2015) US
- (71) **Applicant (for all designated States except US):** INTEL CORPORATION [US/US]; 2200 Mission College Blvd., Santa Clara, California 95054 (US).
- (72) **Inventors; and**
- (71) **Applicants (for US only):** WASSERBLAT, Moshe [IL/IL]; Maccabim Lavi 295, 71799 Maccabim, HA (IL). PEREG, Oren [IL/IL]; PB 401, 37830 Amikam, HA (IL). ASSAYAG, Michel [IL/IL]; 200 Habessor st., 60850 Shoham, M (IL). SIVAK, Alexander [IL/IL]; Hagalil 9A, 32241 Haifa, HA (IL). TAITE, Shahar [IL/IL]; azar 89, 44505 kfar saba, D (IL). RIDER, Tomer [IL/IL]; Wiezman 11, 22407 Naahryia, Z (IL).
- (74) **Agent:** PFLEGER, Edmund; c/o CPA Global, 900 Second Avenue South, Suite 600, Minneapolis, Minnesota 55402 (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, 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, ST, 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))

(54) **Title:** DYNAMIC ADAPTATION OF LANGUAGE MODELS AND SEMANTIC TRACKING FOR AUTOMATIC SPEECH RECOGNITION

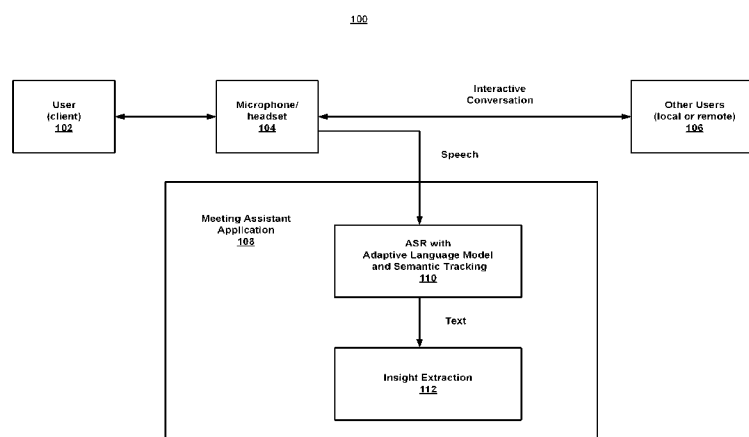


FIG. 1

(57) **Abstract:** Generally, this disclosure provides systems, devices, methods and computer readable media for adaptation of language models and semantic tracking to improve automatic speech recognition (ASR). A system for recognizing phrases of speech from a conversation may include an ASR circuit configured to transcribe a user's speech to a first estimated text sequence, based on a generalized language model. The system may also include a language model matching circuit configured to analyze the first estimated text sequence to determine a context and to select a personalized language model (PLM), from a plurality of PLMs, based on that context. The ASR circuit may further be configured to re-transcribe the speech based on the selected PLM to generate a lattice of paths of estimated text sequences, wherein each of the paths of estimated text sequences comprise one or more words and an acoustic score associated with each of the words.



DYNAMIC ADAPTATION OF LANGUAGE MODELS AND SEMANTIC TRACKING FOR AUTOMATIC SPEECH RECOGNITION

5

10

FIELD

15 The present disclosure relates to automatic speech recognition, and more particularly, to dynamic adaptation of language models and semantic tracking to improve automatic speech recognition.

BACKGROUND

20 Automatic speech recognition (ASR) systems decode a stream of acoustic speech and transcribe it into a sequence of words or text. ASR systems are generally built on classifiers that use a combination of acoustic models and language models to perform that transcription. In order for the ASR system to achieve an improved level of performance, these acoustic and language models must typically be generated from
25 training data that more closely matches the operational environment or scenario in which the ASR will be used. This may include, for example, the speaker's acoustic profile, the context of the conversation and the subject matter or content domain of the conversation. Unfortunately these factors typically vary dynamically over time and existing ASR systems are generally limited to static acoustic and language models,
30 resulting in relatively poorer recognition performance. This is particularly true when trying to recognize conversational speech between humans, as opposed to human-to-machine speech (e.g., speaking to a smartphone digital assistant).

BRIEF DESCRIPTION OF THE DRAWINGS

Features and advantages of embodiments of the claimed subject matter will become apparent as the following Detailed Description proceeds, and upon reference to the Drawings, wherein like numerals depict like parts, and in which:

5 Figure 1 illustrates a top level system diagram of an example embodiment consistent with the present disclosure;

 Figure 2 illustrates a block diagram of one example embodiment consistent with the present disclosure;

 Figure 3 illustrates a block diagram of another example embodiment
10 consistent with the present disclosure;

 Figure 4 illustrates language models consistent with one example embodiment of the present disclosure;

 Figure 5 illustrates language model usage consistent with one example embodiment of the present disclosure;

15 Figure 6 illustrates a sentence lattice consistent with one example embodiment of the present disclosure;

 Figure 7 illustrates a block diagram of another example embodiment consistent with the present disclosure;

 Figure 8 illustrates a flowchart of operations of another example embodiment
20 consistent with the present disclosure; and

 Figure 9 illustrates a system diagram of a platform of another example embodiment consistent with the present disclosure.

 Although the following Detailed Description will proceed with reference being made to illustrative embodiments, many alternatives, modifications, and variations
25 thereof will be apparent to those skilled in the art.

DETAILED DESCRIPTION

Generally, this disclosure provides systems, devices, methods and computer readable media for adaptation of language models and semantic tracking to improve
30 automatic speech recognition (ASR). In some embodiments, a system for recognizing phrases of speech from a conversation may include an ASR circuit configured to transcribe a user's speech to a first estimated text sequence, based on a generalized language model. The system may also include a language model matching circuit

configured to analyze the first estimated text sequence to determine a context and to select a personalized language model (PLM), from a group of available PLMs, based on that context. The ASR circuit may then re-transcribe the speech based on the selected PLM, which may provide improved results over the generalized PLM previously used. The ASR re-transcription may generate a lattice of paths of estimated text sequence possibilities. Each of the paths of estimated text sequences may include one or more words and an acoustic score associated with each of the words.

A final choice selection for the recognized conversational phrase from the lattice may be generated based on a number of factors including the scores, an estimated semantic distance between each lattice path and other previously recognized phrases from the conversation, and estimated contextual relationships between the words in each path, as will be explained in greater detail below.

Figure 1 illustrates a top level system diagram 100 of one example embodiment consistent with the present disclosure. A user (or client) 102 of the system may be engaged in an interactive conversation with other users 106 who may be local or remote. The users may be participants in an online meeting, for example through Skype® or Lync®. The speech of user 102 may be recorded or otherwise sampled through a microphone 104 and provided to an ASR circuit 110 that is configured to provide adaptive language models and semantic tracking for improved recognition, as will be explained in greater detail below. The resulting text transcription of the user's speech, generated by ASR 110, may reach a level of accuracy (e.g., a sufficiently low word error rate) to enable automated insight extraction of the conversation. For example, insight extraction circuit 112 may be configured to generate conversational summaries, extract keywords, detect action requests, estimate sentiments and/or perform language translation, based on the ASR text transcription. In some embodiments, ASR circuit 110 and insight extraction circuit 112 may be incorporated in a meeting assistant application 108 configured to provide a suitable user interface to the users 102 and 106 of the system. A user interface may include, for example, video images of each user, real-time updates of text transcriptions, currently detected insights and various control mechanisms such as joining or leaving, adjusting volume, etc.

Figure 2 illustrates a block diagram of one example embodiment consistent with the present disclosure. The ASR circuit with adaptive language models and semantic tracking 110 is shown in greater detail to include a general language model 206, personalized language models (PLMs) 216, an ASR circuit 208, a language model matching circuit 210, a lattice pruning circuit 212 and a semantic analysis circuit 214, the operations of which will be explained below. Additionally, a PLM generation circuit 204 may be configured to analyze user textual information sources 202 and generate PLMs 216, the operations of which may be performed, for example, in an offline mode (e.g., prior to real-time ASR operation). PLM generation will also be explained in greater detail below.

The ASR circuit 208 is a classifier configured to transcribe speech into text based on acoustic models and language models. The acoustic model defines the speaker's acoustic profile and typically includes statistical representations of various features computed from the speech waveform, for example during a training session. The acoustic model allows the classifier to recognize sounds or phonetic components of words spoken by the user. Language models, in contrast, provide probabilities associated with sequences of words, based on known characteristics of natural language. These probabilities may be used to distinguish between similar sounding words that might otherwise be confused under a given acoustic model. The performance of the ASR (e.g., the word error rate) will generally improve if the acoustic model more closely matches the speaker's voice characteristics and if the language model more closely matches the context in which the user is speaking.

The ASR circuit 208 may be configured to perform an initial transcription of the user's speech to generate a first estimated text sequence based on a generalized language model 206. The text sequence may be a phrase or a sentence spoken by the user. The generalized language model may be designed to match language in a relatively generic context or domain.

Language model matching circuit 210 may be configured to analyze the first estimated text sequence and determine a context based on words found in that text sequence. Language model matching circuit 210 may then select a PLM, from a library or database of previously generated PLMs 216, based on the determined context. The selected PLM may be more closely matched to the current context of the

user's conversational speech and therefore provide improved recognition performance.

The ASR circuit 208 may then re-transcribe the speech using the selected PLM and generate a lattice of text sequence recognition possibilities. An example of a lattice is shown in Figure 6, which will be described in greater detail below. The lattice presents a number of paths out of which a phrase or sentence may be constructed. Each step along the path is associated with one possible word choice along with an associated acoustic score for that word choice provided by the ASR circuit. Lattice pruning circuit 212 may be configured to remove some paths from the lattice, typically paths that involve word choices with relatively lower acoustic scores. In some embodiments, lattice pruning circuit 212 may compare the acoustic scores to a threshold value and disqualify words or entire path segments if the scores fall below the threshold value.

Semantic analysis circuit 214 may be configured to analyze the lattice (or pruned lattice) and to select one of the paths of estimated text sequences as representative of the best choice for the recognized phrase or sentence from the current point in the conversation, as will be described in greater detail below.

Figure 3 illustrates a block diagram of another example embodiment consistent with the present disclosure. PLM generation circuit 204 is shown in greater detail to include information gathering circuit 302, text clustering circuit 308, and knowledge domain generation circuit 310. Information gathering circuit 302 may be configured to search textual information sources associated with the user 202 and, in some embodiments, to search other general knowledge sources 306 or general personal information such as contact lists or address books. User textual information sources may include, for example, electronic documents, emails, text messages, chats, conversational recording and/or social media communications.

Text clustering circuit 308 may be configured to analyze the collected textual information and organize the information into clusters based on a measurement of content similarity between the sources. For example, documents or other sources with relatively high degrees of similarity may be clustered together. One example of a similarity metric may be the number of words that two documents have in common, although any other suitable metric may be used. Knowledge domain generation circuit 310 may be configured to generate domains of knowledge based on each

cluster and to map those domains to various PLMs using any known statistical language modeling techniques or tools. Figure 4 illustrates an example of this process. One cluster of textual sources may be related to a domain of knowledge, DOK #1 402, associated with entertainment media subjects. DOK #1 402 may in turn
5 be mapped to PLM #1 412 that includes words or phrases such as “movie,” “cinema,” “display,” and “television,” etc. Continuing with this example, a second cluster of textual sources may be related to a domain of knowledge, DOK #2 404, associated with photography subjects. DOK #2 404 may be mapped to PLM #2 414 that includes words or phrases such as “photography,” “camera,” and “lens,” etc.
10 Similarly, a third cluster of textual sources may be related to a domain of knowledge, DOK #3 406, associated with computer display hardware subjects. DOK #3 406 may be mapped to PLM #3 416 that includes words or phrases such as “monitor,” “drivers,” and “windows,” etc. This process may continue on for any suitable numbers of DOKs and PLMs.

15

Figure 5 illustrates personalized language model usage 500 consistent with one example embodiment of the present disclosure. Language model matching circuit 210 may be configured to associate text, generated in the first transcription by ASR circuit 208, to a PLM that best matches the current context of the conversation. The
20 text may represent fragments of speech 502, such as, for example, phrases or sentences. In this simplified example, a current snippet of conversation includes three sentences 504, 506, 508:

“I just connected through.”

“I have the graphic on my screen.”

25 “But the resolution is low.”

Sentence 504 represents a previously analyzed sentence at time t-1. Sentence 506 represents the sentence currently being analyzed at time t, and sentence 508 represents the following sentence at time t+1. Based on the context of neighboring sentences, language model matching circuit 210 decides that the current conversational context
30 most likely relates to computer display hardware (for example with a confidence of 56%) and that PLM #3 416 may therefore provide the best results in a second transcription by ASR circuit 208.

Figure 6 illustrates a sentence lattice 600 consistent with one example embodiment of the present disclosure. The second speech transcription by ASR circuit 208, using the PLM selected by language model matching circuit 210, results in a lattice of text sequence recognition possibilities, a simplified example of which is presented in Figure 6. The lattice 600 includes a series of nodes 602, 604, 606, 608, 610 and 612 interconnected by paths associated with word choices and acoustic scores. Many possible phrases or sentences may be derived from this lattice including the following examples:

“... have glasses on my screen ...”

“... have a graphic on my screen ...”

“... and grass on my hand ...”

As previously described, lattice pruning circuit 212 may be configured to remove paths from the lattice with relatively lower acoustic scores. A final selection needs to be made from the remaining paths, however, and this is performed by semantic analysis circuit 214 as described below.

Figure 7 illustrates a block diagram 214 of another example embodiment consistent with the present disclosure. Semantic analysis circuit 214 is shown to include semantic distance calculation circuit 702 and conditional random field (CRF) classifier circuit 704. Continuing with the previous examples of conversational sentences from Figure 5 and the lattice of Figure 6, it would be readily apparent to most humans that the sentence “... have a graphic on my screen ...” is the most likely choice in the lattice based on the current conversational context relating to “connecting to a meeting but experiencing low screen resolution.”

The semantic distance calculation circuit 702 may be configured to perform a similar function by estimating a semantic distance between each of the paths of estimated text sequences to one or more previously recognized phrases of speech from the conversation. In other words, a measure of how well the current sentence can be inferred from an earlier sentence can be made. This may be accomplished using any of a number of known techniques or tools. For example, vector based methods can convert each word or sentence to a vector representation and then semantic similarity can be measured by algebraic vector similarity metrics such as the Cosine Distance metric. The choices with the lowest semantic distance are likely to be the best choices.

Additional processing may be performed by conditional random field (CRF) classifier circuit 704 to further narrow down the lattice possibilities to make the final choice. CRF circuit 704 may be configured to rank each of the paths of estimated text sequences based on contextual relationships between the words in the paths. One
5 example of a relationship that can be exploited for this purpose involves rules of grammar. For example, a noun often follows an adjective (e.g., “nice house”). Of course other any other suitable types of contextual relationship may be used.

The semantic analysis circuit 214 may thus select one of the paths of estimated text sequences, from the lattice, as a currently recognized phrase of speech from the
10 conversation, based on the semantic distance and the CRF ranking.

Figure 8 illustrates a flowchart of operations 800 of another example embodiment consistent with the present disclosure. The operations provide a method for recognizing phrases of speech from a conversation. At operation 810, speech of a
15 participant in the conversation is transcribed to a first estimated text sequence, by an automatic speech recognition (ASR) circuit. The transcription is based on a generalized language model. At operation 820, the first estimated text sequence is analyzed to determine a context. At operation 830, a personalized language model (PLM) is selected from a plurality of PLMs based on the context. At operation 840,
20 the speech is re-transcribed by the ASR circuit, based on the selected PLM, to generate a lattice of paths of estimated text sequences. Each of the paths of estimated text sequences comprise one or more words and an acoustic score associated with each of the words.

25 Figure 9 illustrates a system diagram 900 of one example embodiment consistent with the present disclosure. The system 900 may be a platform 910 hosting a communication and/or computing device such as, for example, a smart phone, smart tablet, personal digital assistant (PDA), mobile Internet device (MID), convertible tablet, notebook or laptop computer, workstation or desktop computer.

30 The system 900 is shown to include one or more processors 920 and memory 930. In some embodiments, the processors 920 may be implemented as any number of processor cores. The processor (or processor cores) may be any type of processor, such as, for example, a micro-processor, an embedded processor, a digital signal processor (DSP), a graphics processor (GPU), a network processor, a field

programmable gate array or other device configured to execute code. The processors may be multithreaded cores in that they may include more than one hardware thread context (or “logical processor”) per core. The memory 930 may be coupled to the processors. The memory 930 may be any of a wide variety of memories (including
5 various layers of memory hierarchy and/or memory caches) as are known or otherwise available to those of skill in the art. It will be appreciated that the processors and memory may be configured to store, host and/or execute one or more operating systems, user applications or other software. The applications may include, but not be limited to, for example, any type of computation, communication, data
10 management, data storage and/or user interface task. In some embodiments, these applications may employ or interact with any other components of the platform 910.

System 900 is also shown to include an ASR circuit 110 configured to provide an adaptive language model and semantic tracking, as described previously.

System 900 is also shown to include network interface circuit 940 which may
15 include wireless communication capabilities, such as, for example, cellular communications, Wireless Fidelity (WiFi), Bluetooth®, and/or Near Field Communication (NFC). The wireless communications may conform to or otherwise be compatible with any existing or yet to be developed communication standards including past, current and future version of Bluetooth®, Wi-Fi and mobile phone
20 communication standards.

System 900 is also shown to include an input/output (IO) system or controller 950 which may be configured to enable or manage data communication between processor 920 and other elements of system 900 or other elements (not shown) external to system 900. The system may generally present various interfaces to a user
25 via a display element 960 such as, for example, a touch screen, liquid crystal display (LCD) or any other suitable display type. Audio may be received through microphone 104, particularly for use by ASR circuit 110. System 900 is also shown to include a storage system 970, for example a hard disk drive (HDD) or solid state drive (SSD), coupled to the processor 920.

30 It will be appreciated that in some embodiments, the various components of the system 900 may be combined in a system-on-a-chip (SoC) architecture. In some embodiments, the components may be hardware components, firmware components, software components or any suitable combination of hardware, firmware or software.

“Circuit” or “circuitry,” as used in any embodiment herein, may comprise, for example, singly or in any combination, hardwired circuitry, programmable circuitry such as computer processors comprising one or more individual instruction processing cores, state machine circuitry, and/or firmware that stores instructions executed by

5 programmable circuitry. The circuitry may include a processor and/or controller configured to execute one or more instructions to perform one or more operations described herein. The instructions may be embodied as, for example, an application, software, firmware, etc. configured to cause the circuitry to perform any of the

10 aforementioned operations. Software may be embodied as a software package, code, instructions, instruction sets and/or data recorded on a computer-readable storage device. Software may be embodied or implemented to include any number of processes, and processes, in turn, may be embodied or implemented to include any number of threads, etc., in a hierarchical fashion. Firmware may be embodied as code, instructions or instruction sets and/or data that are hard-coded (e.g., nonvolatile)

15 in memory devices. The circuitry may, collectively or individually, be embodied as circuitry that forms part of a larger system, for example, an integrated circuit (IC), an application-specific integrated circuit (ASIC), a system on-chip (SoC), desktop computers, laptop computers, tablet computers, servers, smart phones, etc. Other embodiments may be implemented as software executed by a programmable control

20 device. As described herein, various embodiments may be implemented using hardware elements, software elements, or any combination thereof. Examples of hardware elements may include processors, microprocessors, circuits, circuit elements (e.g., transistors, resistors, capacitors, inductors, and so forth), integrated circuits, application specific integrated circuits (ASIC), programmable logic devices (PLD),

25 digital signal processors (DSP), field programmable gate array (FPGA), logic gates, registers, semiconductor device, chips, microchips, chip sets, and so forth.

Any of the operations described herein may be implemented in one or more storage devices having stored thereon, individually or in combination, instructions that when executed by one or more processors perform one or more operations. Also,

30 it is intended that the operations described herein may be performed individually or in any sub-combination. Thus, not all of the operations (for example, of any of the flow charts) need to be performed, and the present disclosure expressly intends that all sub-combinations of such operations are enabled as would be understood by one of ordinary skill in the art. Also, it is intended that operations described herein may be

distributed across a plurality of physical devices, such as processing structures at more than one different physical location. The storage devices may include any type of tangible device, for example, any type of disk including hard disks, floppy disks, optical disks, compact disk read-only memories (CD-ROMs), compact disk rewritables (CD-RWs), and magneto-optical disks, semiconductor devices such as read-only memories (ROMs), random access memories (RAMs) such as dynamic and static RAMs, erasable programmable read-only memories (EPROMs), electrically erasable programmable read-only memories (EEPROMs), flash memories, Solid State Disks (SSDs), magnetic or optical cards, or any type of media suitable for storing electronic instructions.

Thus, the present disclosure provides systems, devices, methods and computer readable media for adaptation of language models and semantic tracking to improve automatic speech recognition (ASR). The following examples pertain to further embodiments.

According to Example 1 there is provided a system for recognizing phrases of speech from a conversation. The system may include an automatic speech recognition (ASR) circuit to transcribe speech, of a user of the system, to a first estimated text sequence, based on a generalized language model; a language model matching circuit to analyze the first estimated text sequence to determine a context and to select a personalized language model (PLM), from a plurality of PLMs, based on the context; and the ASR circuit further configured to re-transcribe the speech based on the selected PLM to generate a lattice of paths of estimated text sequences, and each of the paths of estimated text sequences include one or more words and an acoustic score associated with each of the words.

Example 2 may include the subject matter of Example 1, further including a lattice pruning circuit to remove a subset of the paths of estimated text sequences based on a comparison of the acoustic scores to a threshold value.

Example 3 may include the subject matter of Examples 1 and 2, further including a semantic distance calculation circuit to estimate a semantic distance between each of the paths of estimated text sequences to one or more previously recognized phrases of speech from the conversation.

Example 4 may include the subject matter of Examples 1-3, further including a conditional random field (CRF) classifier circuit to rank each of the paths of estimated text sequences based on contextual relationships between the words in the paths.

Example 5 may include the subject matter of Examples 1-4, further including a semantic analysis circuit to select one of the paths of estimated text sequences, from the lattice, as a currently recognized phrase of speech from the conversation, based on the semantic distance and the CRF ranking.

5 Example 6 may include the subject matter of Examples 1-5, further including a PLM generation circuit to analyze textual information sources associated with the user; to organize the textual information from the sources into clusters based on a measurement of content similarity between the sources, to generate domains of knowledge based on the clusters; and to map the domains into the plurality of PLMs.

10 Example 7 may include the subject matter of Examples 1-6, and the textual information sources include electronic documents, emails, text messages or social media communications.

Example 8 may include the subject matter of Examples 1-7, and the PLM generation circuit operates in an offline mode prior to execution of the ASR circuit.

15 Example 9 may include the subject matter of Examples 1-8, further including an insight extraction circuit to analyze the current and previously recognized phrases of speech from the conversation and to generate a summary of the conversation, extract keywords from the conversation, perform a translation of the conversation or extract action requests taken by the user during the conversation.

20 According to Example 10 there is provided a method for recognizing phrases of speech from a conversation. The method may include: transcribing speech, of a participant in the conversation, to a first estimated text sequence, by an automatic speech recognition (ASR) circuit, the transcription based on a generalized language model; analyzing the first estimated text sequence to determine a context; selecting a
25 personalized language model (PLM), from a plurality of PLMs, based on the context; and re-transcribing the speech, by the ASR circuit, based on the selected PLM, to generate a lattice of paths of estimated text sequences, and each of the paths of estimated text sequences include one or more words and an acoustic score associated with each of the words.

30 Example 11 may include the subject matter of Example 10, further including removing a subset of the paths of estimated text sequences based on a comparison of the acoustic scores to a threshold value.

Example 12 may include the subject matter of Examples 10 and 11, further including estimating a semantic distance between each of the paths of estimated text

sequences to one or more previously recognized phrases of speech from the conversation.

Example 13 may include the subject matter of Examples 10-12, further including determining contextual relationships between the words in the paths and
5 ranking each of the paths of estimated text sequences based on the contextual relationships.

Example 14 may include the subject matter of Examples 10-13, further including selecting one of the paths of estimated text sequences, from the lattice, as a currently recognized phrase of speech from the conversation, based on the semantic
10 distance and the ranking.

Example 15 may include the subject matter of Examples 10-14, further including generating the PLMs by: analyzing textual information sources associated with the participant; organizing the textual information from the sources into clusters based on a measurement of content similarity between the sources; generating
15 domains of knowledge based on the clusters; and mapping the domains into the plurality of PLMs.

Example 16 may include the subject matter of Examples 10-15, and the textual information sources include electronic documents, emails, text messages or social media communications.

20 Example 17 may include the subject matter of Examples 10-16, and the PLM generation is performed in an offline mode prior to execution of the ASR circuit.

Example 18 may include the subject matter of Examples 10-17, further including analyzing the current and previously recognized phrases of speech from the conversation to generate a summary of the conversation, extract keywords from the
25 conversation, perform a translation of the conversation or extract action requests taken by the participant during the conversation.

According to Example 19 there is provided at least one computer-readable storage medium having instructions stored thereon which when executed by a processor result in the following operations for recognizing phrases of speech from a
30 conversation. The operations may include: transcribing speech, of a participant in the conversation, to a first estimated text sequence, by an automatic speech recognition (ASR) circuit, the transcription based on a generalized language model; analyzing the first estimated text sequence to determine a context; selecting a personalized language model (PLM), from a plurality of PLMs, based on the context; and re-transcribing the

speech, by the ASR circuit, based on the selected PLM, to generate a lattice of paths of estimated text sequences, and each of the paths of estimated text sequences include one or more words and an acoustic score associated with each of the words.

Example 20 may include the subject matter of Example 19, further including
5 removing a subset of the paths of estimated text sequences based on a comparison of the acoustic scores to a threshold value.

Example 21 may include the subject matter of Examples 19 and 20, further including estimating a semantic distance between each of the paths of estimated text sequences to one or more previously recognized phrases of speech from the
10 conversation.

Example 22 may include the subject matter of Examples 19-21, further including determining contextual relationships between the words in the paths and ranking each of the paths of estimated text sequences based on the contextual relationships.

Example 23 may include the subject matter of Examples 19-22, further including selecting one of the paths of estimated text sequences, from the lattice, as a currently recognized phrase of speech from the conversation, based on the semantic distance and the ranking.

Example 24 may include the subject matter of Examples 19-23, further
20 including generating the PLMs by: analyzing textual information sources associated with the participant; organizing the textual information from the sources into clusters based on a measurement of content similarity between the sources; generating domains of knowledge based on the clusters; and mapping the domains into the plurality of PLMs.

Example 25 may include the subject matter of Examples 19-24, and the textual information sources include electronic documents, emails, text messages or social media communications.

Example 26 may include the subject matter of Examples 19-25, and the PLM generation is performed in an offline mode prior to execution of the ASR circuit.

Example 27 may include the subject matter of Examples 19-26, further
30 including analyzing the current and previously recognized phrases of speech from the conversation to generate a summary of the conversation, extract keywords from the conversation, perform a translation of the conversation or extract action requests taken by the participant during the conversation.

According to Example 28 there is provided a system for recognizing phrases of speech from a conversation. The system may include: means for transcribing speech, of a participant in the conversation, to a first estimated text sequence, by an automatic speech recognition (ASR) circuit, the transcription based on a generalized language model; means for analyzing the first estimated text sequence to determine a context; means for selecting a personalized language model (PLM), from a plurality of PLMs, based on the context; and means for re-transcribing the speech, by the ASR circuit, based on the selected PLM, to generate a lattice of paths of estimated text sequences, and each of the paths of estimated text sequences include one or more words and an acoustic score associated with each of the words.

Example 29 may include the subject matter of Example 28, further including means for removing a subset of the paths of estimated text sequences based on a comparison of the acoustic scores to a threshold value.

Example 30 may include the subject matter of Examples 28 and 29, further including means for estimating a semantic distance between each of the paths of estimated text sequences to one or more previously recognized phrases of speech from the conversation.

Example 31 may include the subject matter of Examples 28-30, further including means for determining contextual relationships between the words in the paths and ranking each of the paths of estimated text sequences based on the contextual relationships.

Example 32 may include the subject matter of Examples 28-31, further including means for selecting one of the paths of estimated text sequences, from the lattice, as a currently recognized phrase of speech from the conversation, based on the semantic distance and the ranking.

Example 33 may include the subject matter of Examples 28-32, further including means for generating the PLMs by: analyzing textual information sources associated with the participant; organizing the textual information from the sources into clusters based on a measurement of content similarity between the sources; generating domains of knowledge based on the clusters; and mapping the domains into the plurality of PLMs.

Example 34 may include the subject matter of Examples 28-33, and the textual information sources include electronic documents, emails, text messages or social media communications.

Example 35 may include the subject matter of Examples 28-34, and the PLM generation is performed in an offline mode prior to execution of the ASR circuit.

Example 36 may include the subject matter of Examples 28-35, further including means for analyzing the current and previously recognized phrases of
5 speech from the conversation to generate a summary of the conversation, extract keywords from the conversation, perform a translation of the conversation or extract action requests taken by the participant during the conversation.

The terms and expressions which have been employed herein are used as terms of description and not of limitation, and there is no intention, in the use of such
10 terms and expressions, of excluding any equivalents of the features shown and described (or portions thereof), and it is recognized that various modifications are possible within the scope of the claims. Accordingly, the claims are intended to cover all such equivalents. Various features, aspects, and embodiments have been described herein. The features, aspects, and embodiments are susceptible to combination with
15 one another as well as to variation and modification, as will be understood by those having skill in the art. The present disclosure should, therefore, be considered to encompass such combinations, variations, and modifications.

CLAIMS

What is claimed is:

1. A system for recognizing phrases of speech from a conversation, said system comprising:
 - an automatic speech recognition (ASR) circuit to transcribe speech, of a user of said system, to a first estimated text sequence, based on a generalized language model;
 - a language model matching circuit to analyze said first estimated text sequence to determine a context and to select a personalized language model (PLM), from a plurality of PLMs, based on said context; and
 - said ASR circuit further to re-transcribe said speech based on said selected PLM to generate a lattice of paths of estimated text sequences, wherein each of said paths of estimated text sequences comprise one or more words and an acoustic score associated with each of said words.
2. The system of claim 1, further comprising a lattice pruning circuit to remove a subset of said paths of estimated text sequences based on a comparison of said acoustic scores to a threshold value.
3. The system of claim 1, further comprising a semantic distance calculation circuit to estimate a semantic distance between each of said paths of estimated text sequences to one or more previously recognized phrases of speech from said conversation.
4. The system of claim 3, further comprising a conditional random field (CRF) classifier circuit to rank each of said paths of estimated text sequences based on contextual relationships between said words in said paths.
5. The system of claim 4, further comprising a semantic analysis circuit to select one of said paths of estimated text sequences, from said lattice, as a currently recognized phrase of speech from said conversation, based on said semantic distance and said CRF ranking.

6. The system of claim 1, further comprising a PLM generation circuit to analyze textual information sources associated with said user; to organize said textual information from said sources into clusters based on a measurement of content similarity between said sources, to generate domains of knowledge based on said clusters; and to map said domains into said plurality of PLMs.

7. The system of claim 6, wherein said textual information sources comprise electronic documents, emails, text messages or social media communications.

8. The system of claim 6, wherein said PLM generation circuit operates in an offline mode prior to execution of said ASR circuit.

9. The system of claim 5, further comprising an insight extraction circuit to analyze said current and previously recognized phrases of speech from said conversation and to generate a summary of said conversation, extract keywords from said conversation, perform a translation of said conversation or extract action requests taken by said user during said conversation.

10. A method for recognizing phrases of speech from a conversation, said method comprising:

transcribing speech, of a participant in said conversation, to a first estimated text sequence, by an automatic speech recognition (ASR) circuit, said transcription based on a generalized language model;

analyzing said first estimated text sequence to determine a context;

selecting a personalized language model (PLM), from a plurality of PLMs, based on said context; and

re-transcribing said speech, by said ASR circuit, based on said selected PLM, to generate a lattice of paths of estimated text sequences, wherein each of said paths of estimated text sequences comprise one or more words and an acoustic score associated with each of said words.

11. The method of claim 10, further comprising removing a subset of said paths of estimated text sequences based on a comparison of said acoustic scores to a threshold value.

12. The method of claim 10, further comprising estimating a semantic distance between each of said paths of estimated text sequences to one or more previously recognized phrases of speech from said conversation.
13. The method of claim 12, further comprising determining contextual relationships between said words in said paths and ranking each of said paths of estimated text sequences based on said contextual relationships.
14. The method of claim 13, further comprising selecting one of said paths of estimated text sequences, from said lattice, as a currently recognized phrase of speech from said conversation, based on said semantic distance and said ranking.
15. The method of claim 10, further comprising generating said PLMs by:
 - analyzing textual information sources associated with said participant;
 - organizing said textual information from said sources into clusters based on a measurement of content similarity between said sources;
 - generating domains of knowledge based on said clusters; and
 - mapping said domains into said plurality of PLMs.
16. The method of claim 15, wherein said textual information sources comprise electronic documents, emails, text messages or social media communications.
17. The method of claim 15, wherein said PLM generation is performed in an offline mode prior to execution of said ASR circuit.
18. The method of claim 14, further comprising analyzing said current and previously recognized phrases of speech from said conversation to generate a summary of said conversation, extract keywords from said conversation, perform a translation of said conversation or extract action requests taken by said participant during said conversation.

19. At least one computer-readable storage medium having instructions stored thereon which when executed by a processor result in the following operations for recognizing phrases of speech from a conversation, said operations comprising:
- transcribing speech, of a participant in said conversation, to a first estimated text sequence, by an automatic speech recognition (ASR) circuit, said transcription based on a generalized language model;
 - analyzing said first estimated text sequence to determine a context;
 - selecting a personalized language model (PLM), from a plurality of PLMs, based on said context; and
 - re-transcribing said speech, by said ASR circuit, based on said selected PLM, to generate a lattice of paths of estimated text sequences, wherein each of said paths of estimated text sequences comprise one or more words and an acoustic score associated with each of said words.
20. The computer-readable storage medium of claim 19, further comprising removing a subset of said paths of estimated text sequences based on a comparison of said acoustic scores to a threshold value.
21. The computer-readable storage medium of claim 19, further comprising estimating a semantic distance between each of said paths of estimated text sequences to one or more previously recognized phrases of speech from said conversation.
22. The computer-readable storage medium of claim 21, further comprising determining contextual relationships between said words in said paths and ranking each of said paths of estimated text sequences based on said contextual relationships.
23. The computer-readable storage medium of claim 22, further comprising selecting one of said paths of estimated text sequences, from said lattice, as a currently recognized phrase of speech from said conversation, based on said semantic distance and said ranking.
24. The computer-readable storage medium of claim 19, further comprising generating said PLMs by:
- analyzing textual information sources associated with said participant;

organizing said textual information from said sources into clusters based on a measurement of content similarity between said sources;
generating domains of knowledge based on said clusters; and
mapping said domains into said plurality of PLMs.

25. The computer-readable storage medium of claim 24, wherein said textual information sources comprise electronic documents, emails, text messages or social media communications.

26. The computer-readable storage medium of claim 24, wherein said PLM generation is performed in an offline mode prior to execution of said ASR circuit.

27. The computer-readable storage medium of claim 23, further comprising analyzing said current and previously recognized phrases of speech from said conversation to generate a summary of said conversation, extract keywords from said conversation, perform a translation of said conversation or extract action requests taken by said participant during said conversation.

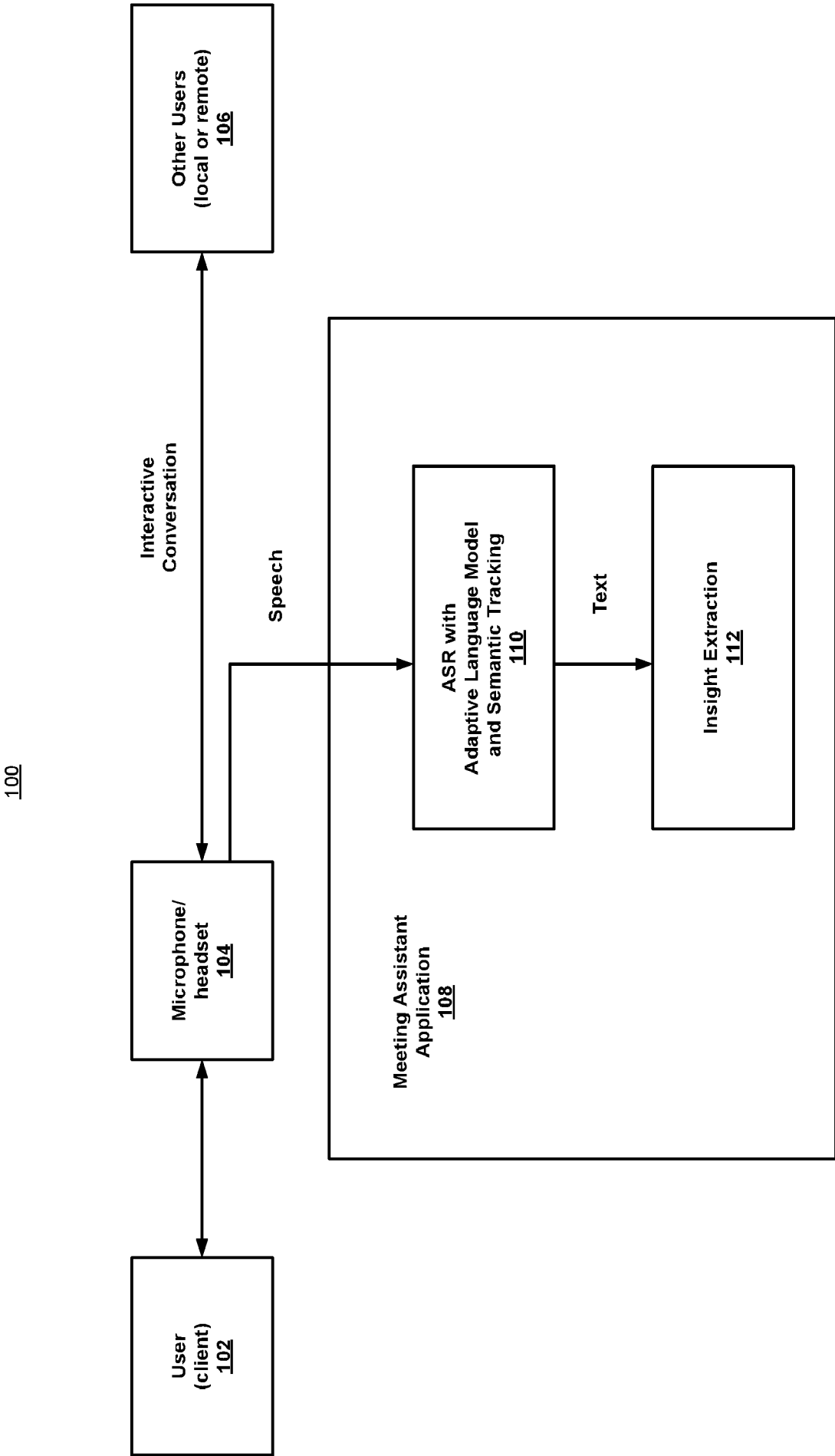


FIG. 1

110

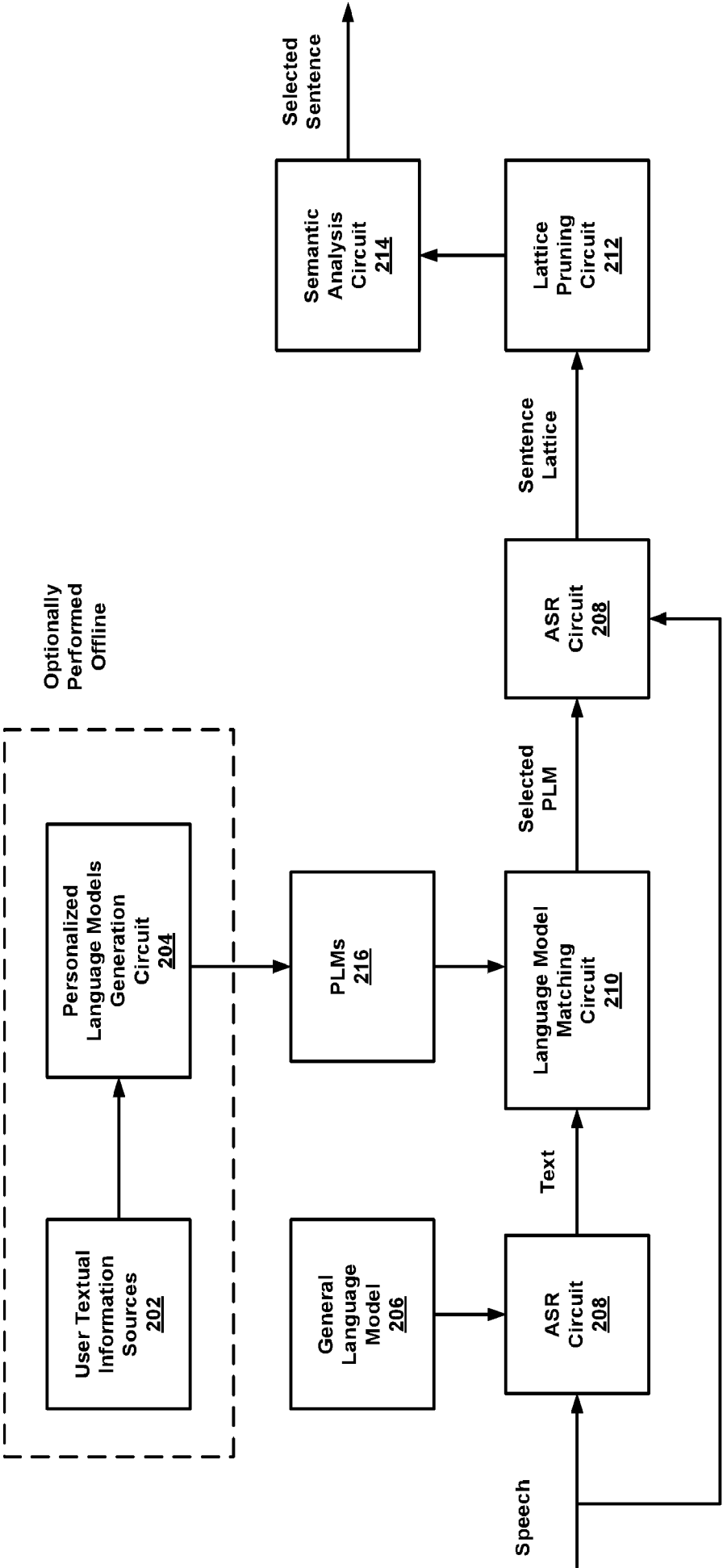


FIG. 2

204

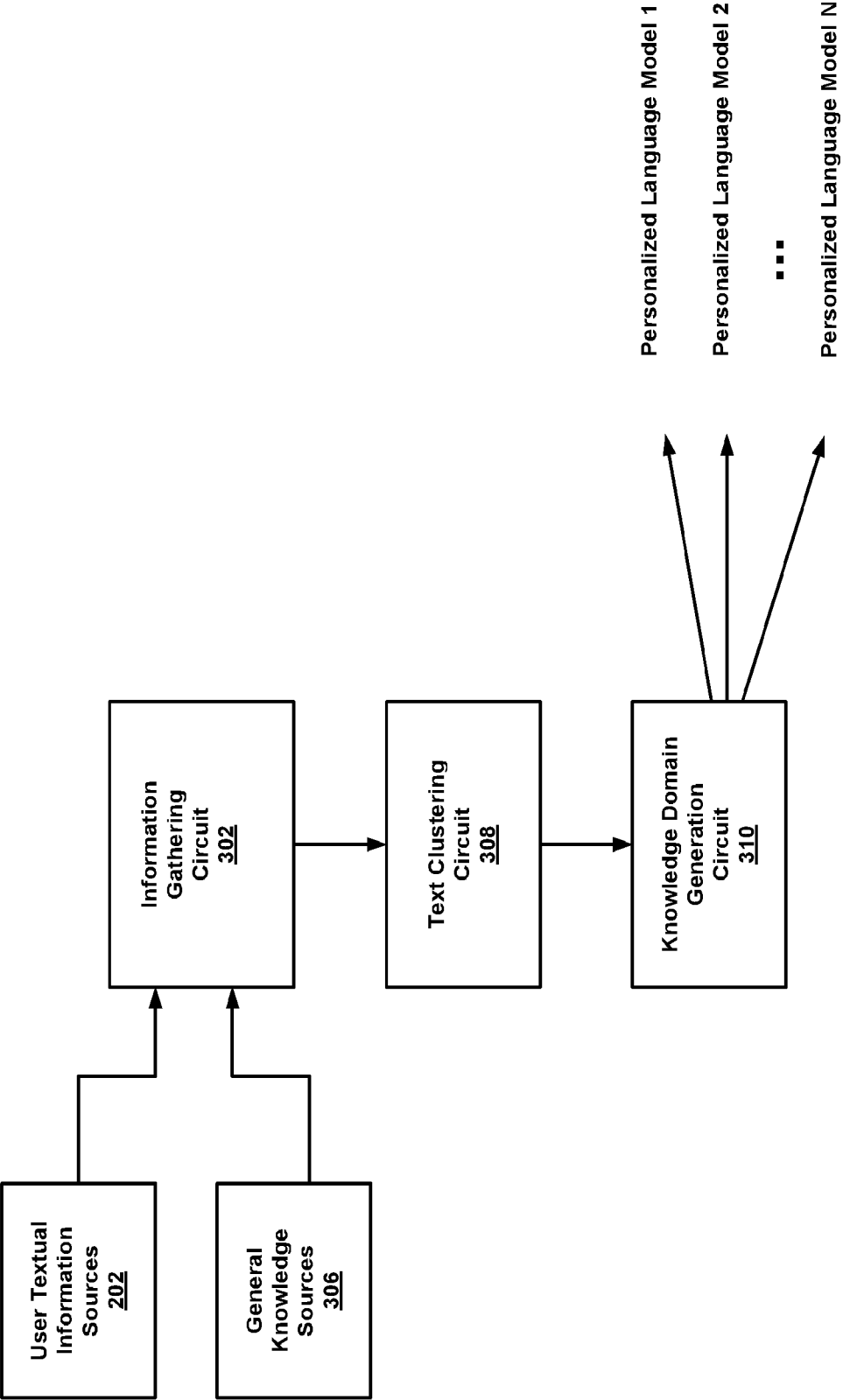


FIG. 3

400

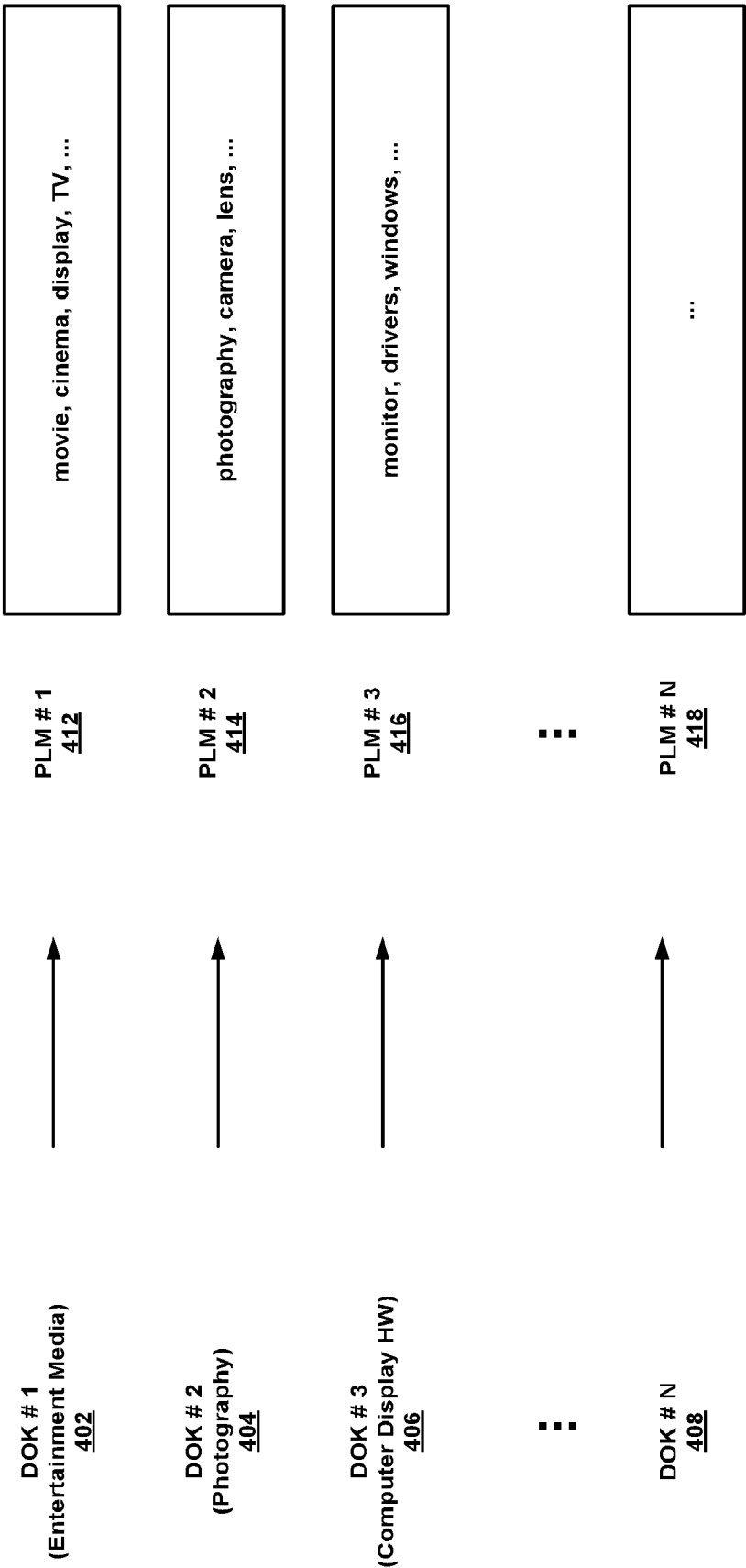


FIG. 4

500

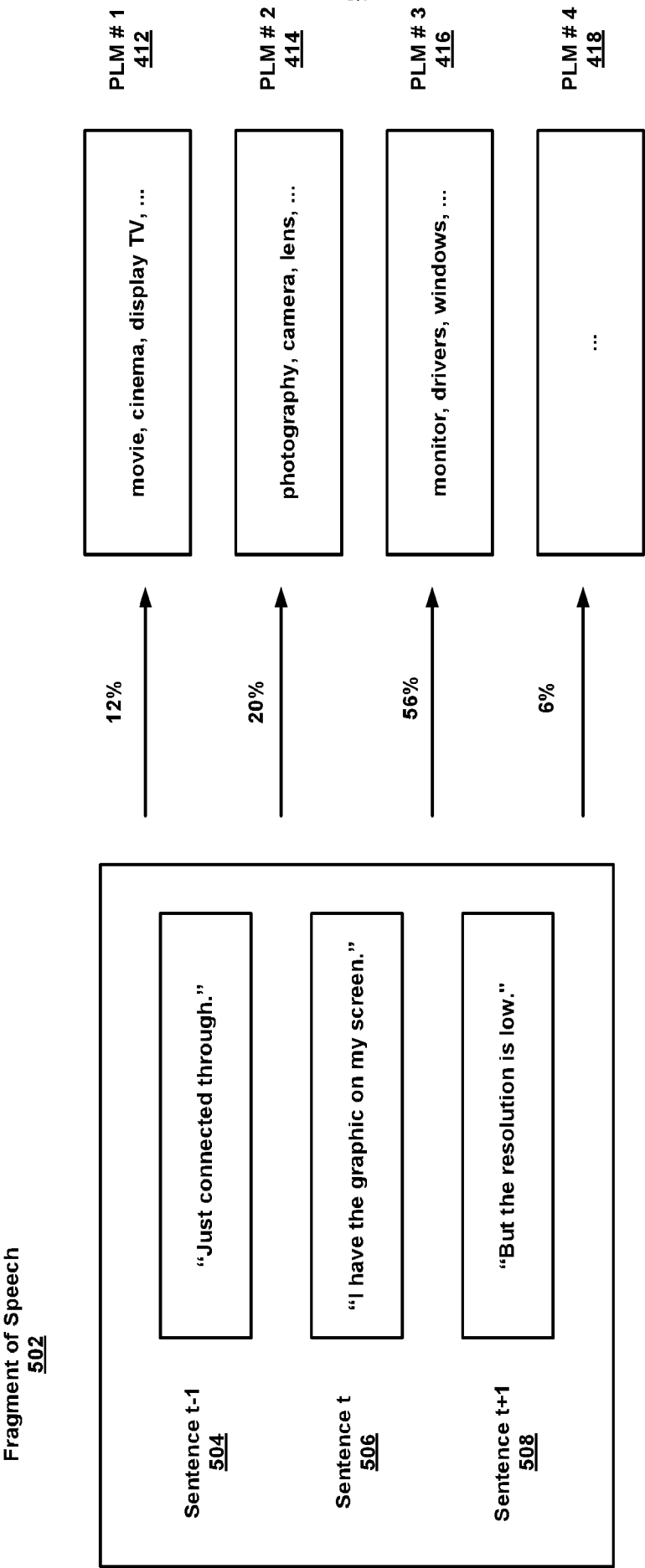
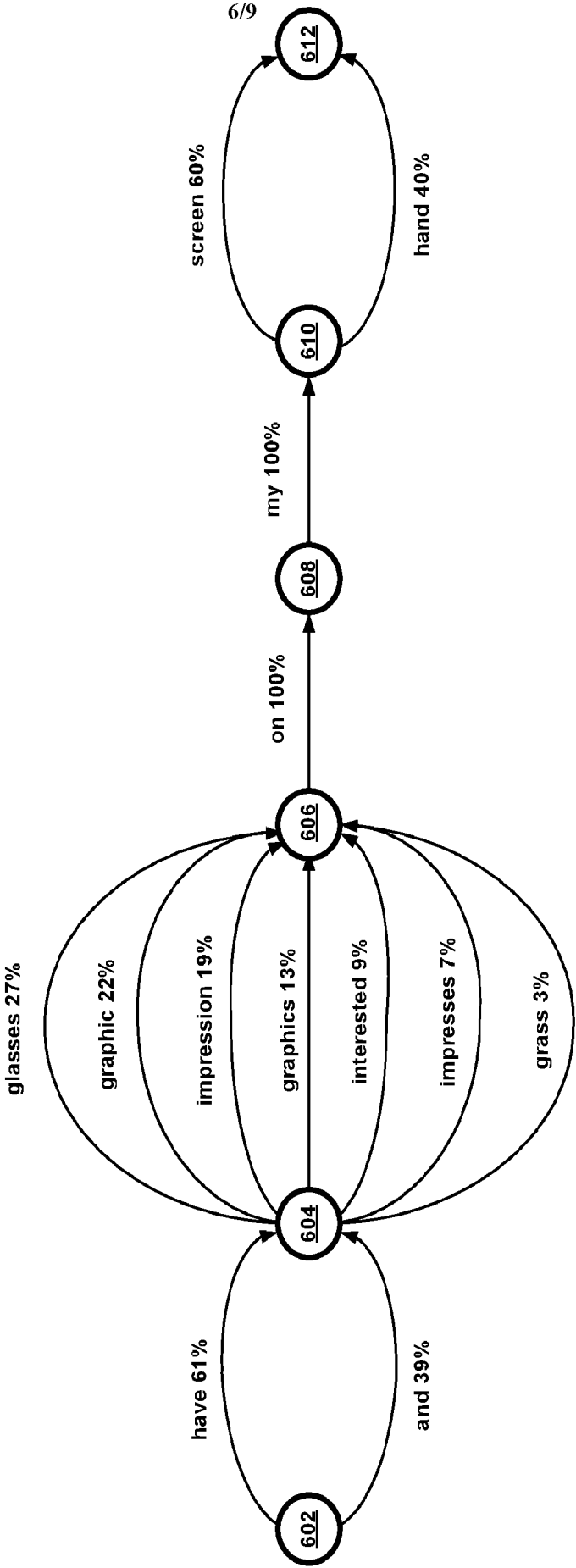


FIG. 5

Sentence
Lattice
600



6/9

FIG. 6

214

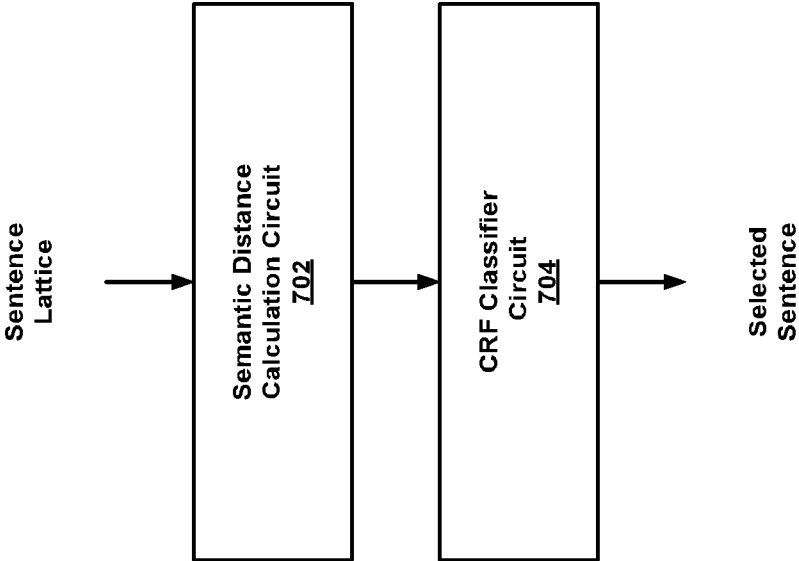


FIG. 7

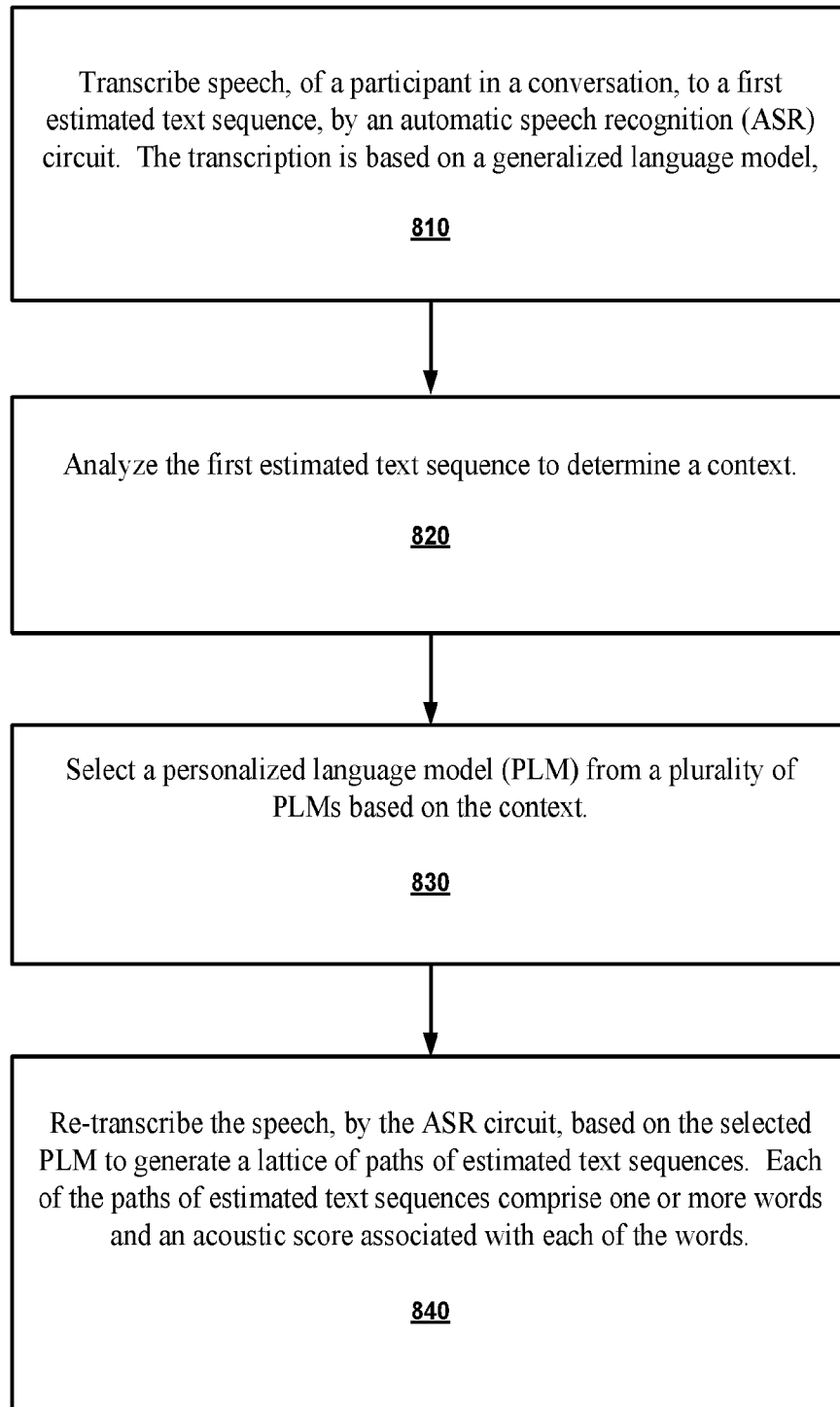


FIG. 8

900

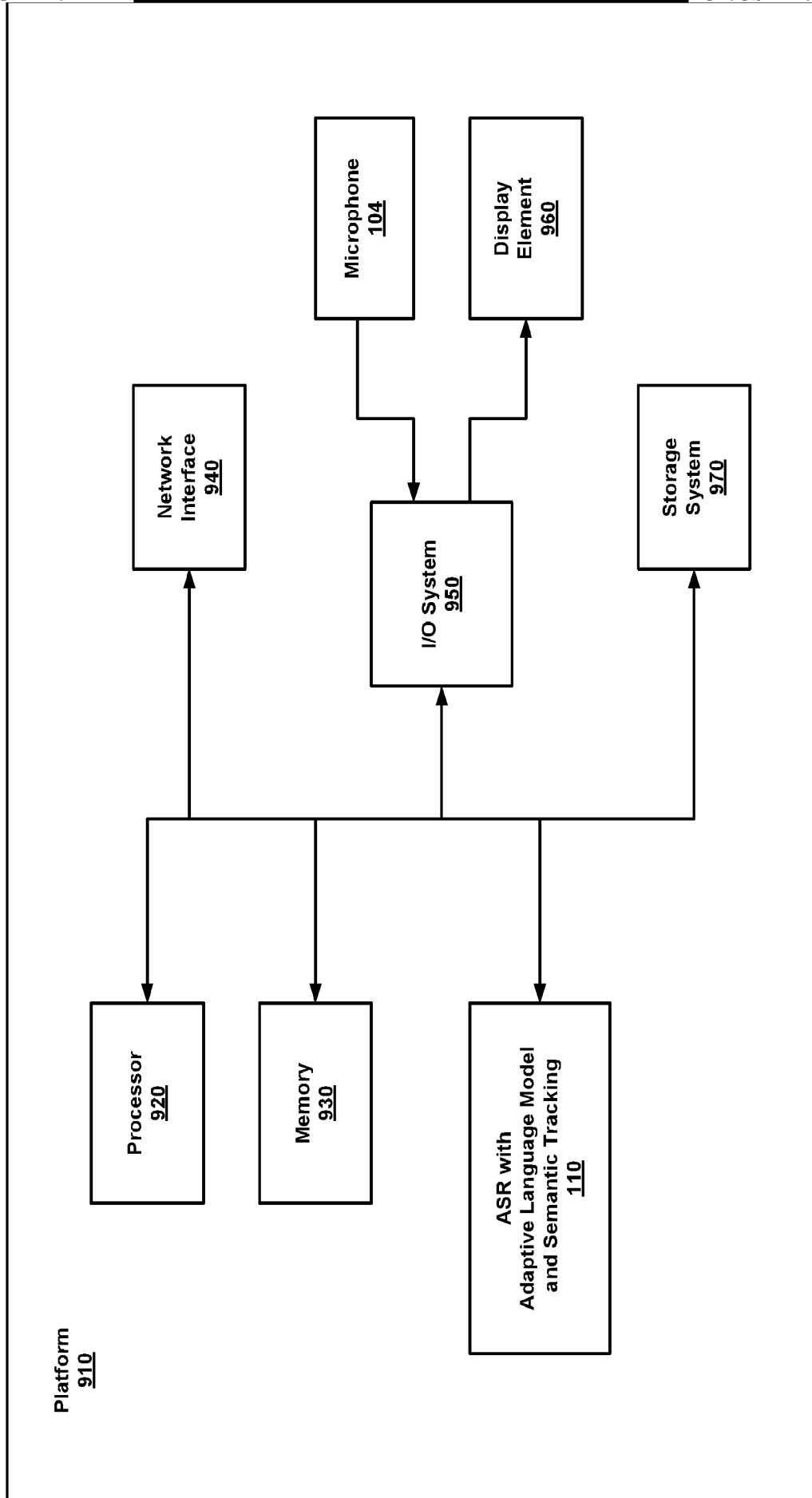


FIG. 9

A. CLASSIFICATION OF SUBJECT MATTER**G10L 15/14(2006.01)i, G10L 15/30(2013.01)i, G10L 15/183(2013.01)i**

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)

G10L 15/14; G10L 15/06; G10L 13/08; G10L 15/26; G10L 15/22; G06F 17/27; G10L 15/18; G10L 15/30; G10L 15/183

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: speech recognition, general model, personalized language model, context, lattice, semantic, and similar terms.**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 2506252 A2 (NUANCE COMMUNICATIONS AUSTRIA GMBH) 03 October 2012 See paragraphs [0007], [0010]–[0011], [0040]–[0043], [0045], and [0048]; claim 9; and figures 2–4.	1–27
Y	US 2007-0106512 A1 (ALEJANDRO ACERO et al.) 10 May 2007 See paragraphs [0005], [0009], and [0045]; and figure 4.	1–27
Y	US 2014-0257794 A1 (NUANCE COMMUNICATIONS, INC.) 11 September 2014 See paragraphs [0040]–[0048] and figures 4–5.	3–5, 9, 12–14, 18 , 21–23, 27
A	US 2013-0262106 A1 (EYAL HURVITZ et al.) 03 October 2013 See paragraphs [0026]–[0030], [0037]–[0038], [0040], and [0042]–[0051]; and figures 1–2.	1–27
A	US 2010-0023320 A1 (PHILIPPE DI CRISTO et al.) 28 January 2010 See paragraphs [0014], [0016]–[0018], [0099]–[0107], and [0115]–[0118]; and figures 7–8.	1–27



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

17 November 2016 (17.11.2016)

Date of mailing of the international search report

17 November 2016 (17.11.2016)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2016/046162

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2506252 A2	03/10/2012	EP 1687807 A2 EP 1687807 B1 EP 2506252 A3 JP 2007-512608 A JP 2011-186491 A JP 5255769 B2 JP 5330450 B2 US 2007-0271086 A1 US 8041566 B2 WO 2005-050621 A2 WO 2005-050621 A3	09/08/2006 16/03/2016 28/11/2012 17/05/2007 22/09/2011 07/08/2013 30/10/2013 22/11/2007 18/10/2011 02/06/2005 27/10/2005
US 2007-0106512 A1	10/05/2007	CN 101305362 A CN 101305362 B EP 1949260 A1 EP 1949260 A4 KR 10-2008-0069990 A US 7831428 B2 WO 2007-056029 A1	12/11/2008 18/04/2012 30/07/2008 20/07/2011 29/07/2008 09/11/2010 18/05/2007
US 2014-0257794 A1	11/09/2014	US 9269354 B2	23/02/2016
US 2013-0262106 A1	03/10/2013	US 8996371 B2	31/03/2015
US 2010-0023320 A1	28/01/2010	US 2007-0038436 A1 US 2011-0131036 A1 US 2014-0108013 A1 US 7620549 B2 US 8332224 B2 US 8620659 B2 WO 2007-021587 A2 WO 2007-021587 A3	15/02/2007 02/06/2011 17/04/2014 17/11/2009 11/12/2012 31/12/2013 22/02/2007 27/09/2007