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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2007 (18.10.2007)

PCT

(10) International Publication Number
WO 2007/118032 A2

(51) International Patent Classification:
G10L 15/06 (2006.01)

(21) International Application Number:
PCT/US2007/065652

(22) International Filing Date: 30 March 2007 (30.03.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/788,621	3 April 2006 (03.04.2006)	US
60/788,622	3 April 2006 (03.04.2006)	US
60/788,606	3 April 2006 (03.04.2006)	US
11/539,456	6 October 2006 (06.10.2006)	US
11/688,920	21 March 2007 (21.03.2007)	US
11/688,916	21 March 2007 (21.03.2007)	US

(71) Applicant *(for all designated States except US)*: **VOCOLLECT, INC.** [US/US]; 703 Rodi Road, Pittsburgh, PA 15235-4558 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BRAHO, Keith** [US/US]; 3943 Murry Lane, Murrysville, PA 15668 (US). **PIKE, Jeffrey** [US/US]; 5070 Harvest Drive, Murrysville, PA 15668 (US). **PIKE, Lori, A.** [US/US]; 5070 Harvest Drive, Murrysville, PA 15668 (US).

(74) Agents: **TITTEL, Bruce** et al.; Wood, Herron & Evans, L.L.P., 441 Vine Street, 2700 Carew Tower, Cincinnati, OH 45202 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

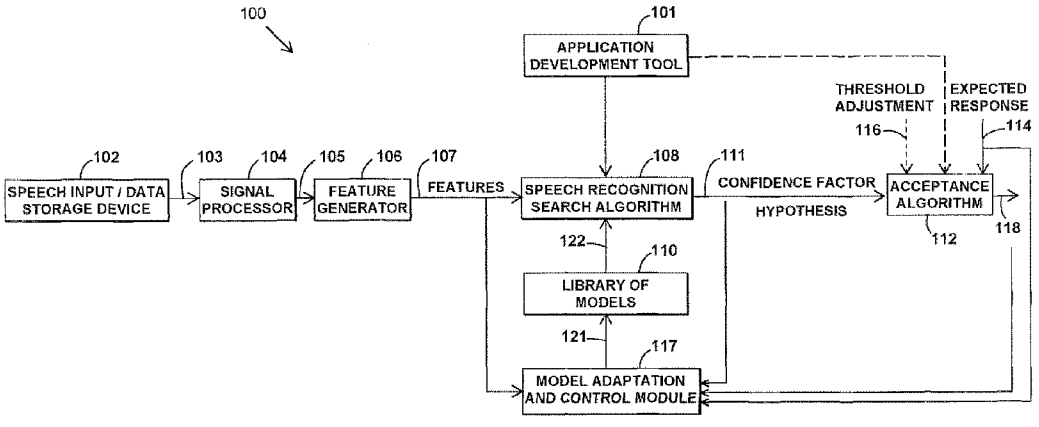
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHODS AND SYSTEMS FOR ADAPTING A MODEL FOR A SPEECH RECOGNITION SYSTEM



(57) **Abstract:** Methods are disclosed for identifying possible errors made by a speech recognition system without using a transcript of words input to the system. A method for model adaptation for a speech recognition system includes determining an error rate (315), corresponding to either recognition of instances of a word or recognition of instances of various words, without using a transcript of words input to the system. The method may further include adjusting an adaptation 320, of the model for the word or various models for the various words, based on the error rate. Apparatus are disclosed for identifying possible errors made by a speech recognition system without using a transcript of words input to the system. An apparatus for model adaptation for a speech recognition system includes a processor 210 adapted to estimate an error rate, corresponding to either recognition of instances of a word or recognition of instances of various words, without using a transcript of words input to the system. The apparatus may further include a controller 225 adapted to adjust an adaptation of the model 235 for the word or various models for the various words, based on the error rate.

WO 2007/118032 A2

METHODS AND SYSTEMS FOR ADAPTING A MODEL FOR A SPEECH RECOGNITION SYSTEM

RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application No. 11/331,649, entitled "Methods and Systems for Considering Information About an Expected Response When Performing Speech Recognition" filed January 13, 2006, which is a continuation-in-part application of U.S. Patent Application No. 11/051,825, entitled "Method and System for Considering Information About an Expected Response When Performing Speech Recognition" filed February 4, 2005, which all applications are incorporated herein by reference in entirety. This application also claims the benefit of U.S. Provisional Application No. 60/788,621, entitled "Methods and Systems for Adapting a Model for a Speech Recognition System", filed April 3, 2006, U.S. Provisional Application No. 60/788,606, entitled "Methods and Systems for Optimizing Model Adaptation for a Speech Recognition System", filed April 3, 2006, and U.S. Provisional Application No. 60/788,622, entitled "Method and Systems for Assessing and Improving the Performance of a Speech Recognition System", filed April 3, 2006, which all applications are incorporated herein by reference in entirety.

TECHNICAL FIELD

The invention relates to speech recognition and, more particularly, to improving the accuracy and efficiency of speech recognition systems.

BACKGROUND

Speech recognition systems have simplified many tasks particularly for a user in the workplace by permitting the user to perform hands-free communication with a computer as a convenient alternative to communication via conventional peripheral input/output devices. For example, a user could wear a wireless wearable terminal having a speech recognition system that permits communication between the user and a central computer system so that the user can receive work assignments and instructions from the central computer system. The user could also communicate to the central computer system information such as data entries, questions, work progress reports, and work condition reports. In a warehouse or inventory environment, a user can be directed (through an audio instruction from the central computer system or visually by means of a display) to a particular work area that is labeled with a multiple-digit number (check-digit) such as "1-2-3" and be asked to speak the check-digit. The user would then respond with the expected response "1-2-3". (Note that a "check-digit" can be any word or sequence of words, and is not limited to digits.)

Other such examples of applications and communications where knowledge about the response is known are described in U.S. Patent Application No. 2003/0154075 and include environments where a wearable or portable terminal is not required such as in an automobile or a telephone system; environments that are not in a warehouse such as in a pharmacy, retail store, and office; voice-controlled information processing systems that process for example credit card numbers, bank account numbers, social security numbers and personal identification numbers; other applications such

as command and control, dictation, data entry and information retrieval applications; and speech recognition system features such as user verification, password verification, quantity verification, and repeat/acknowledge messages. The inventions presented here can be used in those applications. In using a speech recognition system, manual data entry is eliminated or at the least reduced, and users can perform their tasks faster, more accurately, and more productively.

Errors can be made by a speech recognition system however, due to for example background noise or a user's unfamiliarity or misuse of the system. The errors made by a system can be classified into various types. A metric, the word error rate (which can be defined as the percentage or ratio of speech recognition errors over the number of words input to the system and which can be determined over a window of time and/or data and per user) is often used to evaluate the number and types of errors made by a speech recognition system and is thus useful in evaluating the performance of the system. In general, a word error rate can be determined for a word or for various words among a set of words, or for a user or multiple users. Identification of a system's errors can be done by comparing a reference transcription of a user's input speech to the hypothesis generated by the system (the system's interpretation of the user's input speech). Furthermore, as known to those skilled in the art, the comparison can be performed in a time-aligned mode or in a text-aligned mode.

One type of speech recognition error is a substitution, in which the speech recognition system's hypothesis replaces a word that is in the reference transcription with an incorrect word. For example, if system

recognizes "1-5-3" in response to the user's input speech "1-2-3", the system made one substitution: substituting the '5' for the '2'.

Another type of speech recognition error is a deletion, in which the speech recognition system's hypothesis lacks a word that is in the reference transcription. For example, if system recognizes "1-3" in response to the user's input speech "1-2-3", the system deleted one word, the '2'. There are many types of deletion errors. One variation of the deletion error is a deletion due to recognizing garbage, in which the system erroneously recognizes a garbage model instead of recognizing an actual word. Another variation of the deletion error is a deletion due to a speech misdetection, where the system fails to detect that the audio input to the system contains speech and as a result does not submit features of the audio input to the system's search algorithm. Another type of deletion occurs when the system rejects a correct recognition due to a low confidence score. Yet another variation of the deletion error is a deletion due to a rejected substitution, where a search algorithm of the speech recognition generates a substitution which is later rejected by an acceptance algorithm of the system. Still another type of deletion, occurring in time-aligned comparisons, is a merge: the speech recognition system recognizes two spoken words as one. For example, the user says "four two" and the system outputs "forty".

In this application, a garbage model refers to the general class of models for sounds that do not convey information. Examples may include for example models of breath noises, "um", "uh", sniffles, wind noise, the sound

of a pallet dropping, the sound of a car door slamming, or other general model such as a wildcard. (A wildcard is intended to match the input audio for any audio that doesn't match a model in the library of models.)

Yet another type of speech recognition error is an insertion, in which the speech recognition system's hypothesis includes a word (or symbol) that does not correspond to any word in the reference transcription. Insertion errors often occur when the system generates two symbols that correspond to one symbol. One of these symbols may correspond to the reference transcription and be tagged as a correct recognition. If it does not correspond to the reference transcription, it can be tagged as a substitution error. In either case, the other symbol can be tagged as an insertion error. Insertion errors are also common when noise is mistakenly recognized as speech.

In contrast to determining that an actual error occurred by comparing a system's hypothesis to words actually spoken in a reference transcript, an error can be estimated or deemed to have occurred based on system behavior and user behavior. Accordingly, one can estimate or evaluate the performance level of the speech recognition system, by detecting in this manner the various errors committed by the system. One way to detect a speech recognition error is based on feedback a user provides to the speech recognition system. Feedback can be requested by the speech recognition system. For example, the system could ask the user to confirm the system's hypothesis by asking the user for example "Did you say 1-5-3?", and if the user responds "no", it indicates that the system made an error recognizing "1-5-3". Another type of feedback is based on a user's emotion detected by speech recognition. For example, if the system recognizes in the user's input

speech that the user is sighing or saying words indicating aggravation, it may indicate that an error occurred. Yet another type of feedback is based on a user's correction command to the system, such as the user speaking "back-up" or "erase", or the user identifying what word was spoken (which could be from a list of possible words displayed by the system). When a correction is commanded to the system, it may indicate that an error occurred.

A speech recognition system can improve its performance over time, as more speech samples are received and processed by a speech recognition system, by improving its acoustic models through training or other learning or adaptation algorithms. At the same time, it is useful to prevent the system from adapting in an undesirable way, thereby resulting in a system that performs worse than it did prior to adaptation or a system that degrades over time. Avoiding additional processing by a speech recognition system due to adaptation of acoustic models is particularly useful in many applications, particularly those employing a battery powered mobile computer, wireless network, and server to store models. Adapting models can use significant computational resources to create the adapted models and radio transmission energy to transmit the new models to the server. Example embodiments of the invention disclosed herein can control the rate of adaptation of the speech recognition system to avoid inefficient use of computational, storage and/or power resources and to avoid adapting away from well-performing models. Example embodiments of the invention control adaptation by using triggers, which are based on an error rate determination

(which may be based on an error rate estimation), to cause the adaptation of prior models or create new models. The invention also discloses methods by which recognition error rates can be estimated.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate speech recognition system components and embodiments of the invention and, together with the detailed description of the embodiments given below, serve to explain the principles of the invention.

FIG. 1A illustrates a perspective view of a user using a portable terminal and headset, according to an example embodiment of the invention;

FIG. 1B illustrates a schematic view of a speech recognition system, according to an example embodiment of the invention;

FIG. 2 illustrates a schematic view of a component of a speech recognition system, according to an example embodiment of the invention;

FIG. 3 is a flowchart illustrating a method for controlling model adaptation based on an error rate determination or estimation, according to an example embodiment of the invention;

FIGs. 4-6 are flowcharts illustrating methods for estimating an error rate, according to example embodiments of the invention; and

FIG. 7 is a flowchart illustrating a method for model adaptation, according to an example embodiment of the invention.

DETAILED DESCRIPTION OF EXAMPLE EMBODIMENTS OF THE INVENTION

Example embodiments of the invention disclosed herein control the rate of adaptation of the speech recognition system, using triggers to cause adaptation of prior models or create new models. As a result, example embodiments avoid inefficient use of system resources and avoid adapting away from well-performing models. Example embodiments of the invention include various error rate determinations (which may be based on error rate estimations) which can be used as triggers for model adaptation. Note that in this description, references to "one embodiment" or "an embodiment" mean that the feature being referred to is included in at least one embodiment of the invention. Further, separate references to "one embodiment" in this description do not necessarily refer to the same embodiment; however, neither are such embodiments mutually exclusive, unless so stated and except as will be readily apparent to those skilled in the art. Thus, the invention can include any variety of combinations and/or integrations of the embodiments described herein.

In one approach, a method for efficient use of model adaptation resources of a speech recognition system includes determining an error rate, corresponding to either recognition of instances of a word or recognition of instances of various words among a set of words. The method may further include adjusting an adaptation of a model for the word or various models for the various words, based on the error rate. The approach may be implemented in an apparatus which may include all or a subset of the following: a processor adapted to determine an error rate, corresponding to

either recognition of instances of a word or recognition of instances of various words among a set of words; and a controller adapted to adjust an adaptation of a model for the word or various models for the various words, based on the error rate.

In another approach, a method for identifying a possible error made by a speech recognition system includes identifying an instance of a word that was recognized by the system within a certain confidence factor range. The approach may be implemented as an apparatus which includes a processor adapted to identify an instance of a word that was recognized by the system within a certain confidence factor range.

Yet in another approach, a method for identifying a possible error made by a speech recognition system includes identifying an instance where the system rejects a first hypothesis of a first utterance, followed by the system accepting a second hypothesis of a second utterance, wherein the first and second hypotheses substantially match word-for-word. The approach may be implemented as an apparatus which includes a processor adapted to identify an instance where the system rejects a first hypothesis of a first utterance, followed by the system accepting a second hypothesis of a second utterance, wherein the first and second hypotheses substantially match word-for-word.

In yet another approach, a method for identifying a possible error made by a speech recognition system includes identifying when the system generates a first hypothesis of a first utterance and a second hypothesis of a second utterance and the system accepts the second hypothesis, wherein the two hypotheses do not match word-for-word, but the hypotheses mostly

match word-for-word. The approach may be implemented as an apparatus which includes a processor adapted to identify when the system generates a first hypothesis of a first utterance and a second hypothesis of a second utterance and the system accepts the second hypothesis, wherein the two hypotheses do not match word-for-word, but the hypotheses mostly match word-for-word.

In another approach, a method for identifying a possible error made by a speech recognition system includes identifying when a hypothesis generated by the system does not match an expected response word-for-word, but the hypothesis mostly matches the expected response word-for-word. The approach may be implemented as an apparatus which includes a processor adapted identify when a hypothesis generated by the system does not match an expected response word-for-word, but the hypothesis mostly matches the expected response word-for-word.

Still in another approach, a method for adapting a model for a speech recognition system includes generating a count of occurrences of when a user provides feedback to the system. The method may further include adjusting adaptation of the model based on the count. The approach may be implemented as an apparatus which may include all or a subset of the following: a processor adapted to generate a count of occurrences of when a user provides feedback to the system; and a controller that adjusts an adaptation of the model based on the count.

Example Embodiments of Speech Recognition Systems

FIG. 1A illustrates a perspective view of a user using a portable terminal 10 and headset 16, according to an example embodiment of the

invention. Portable terminal 10 may be a wearable device, which may be worn by a user 11 such as on a belt 14 as shown. Use of the term "portable terminal" herein is not limited and may include any computer, device, machine, or system which is used to perform a specific task. Portable terminal 10 may comprise processing circuitry, including a processor for controlling the operation of the portable terminal and other associate processing circuitry. (The processing circuitry may implement one or more exemplary embodiment speech recognition methods disclosed herein.) Headset 16 may be coupled to the portable terminal by a cord 18 or by a wireless connection (not shown in FIG. 1A) and the headset is worn on the head of the user 11. (In another exemplary embodiment, the speech recognition system is located in headset 16, eliminating the need for portable terminal 10.) A user can speak in a spoken language, through a microphone in the headset 16 and the audio information is processed by the speech recognition system in portable terminal 10. U.S. Patent Application Serial No. 10/671,142, entitled "Apparatus and Method for Detecting User Speech", incorporated herein by reference, provides further details for implementing such a system. Applications for example embodiments of the invention are not strictly limited to the warehouse environment and portable terminals 10 shown in FIG. 1A. Other applications and environments in which example embodiments may be implemented are described in the BACKGROUND section of this application.

FIG. 1B illustrates a schematic view of a speech recognition system, according to an example embodiment of the invention. One of ordinary skill in the art will recognize that the various functional blocks of the speech

recognition system, including the inventive features, can be implemented using a variety of different technologies. The invention can be implemented into various different hardware and software configurations. One particular application for the invention is within a portable or wearable terminal that is used for voice-directed work. However, other implementations are possible as well and the invention is not limited to such voice-directed work applications. Furthermore, to implement various features of the inventive embodiment, the speech recognition software might be modified. Alternatively, a separate module might be utilized to modify the speech recognition system according to aspects of the invention.

Referring to FIG. 1B in system 100, a speech signal, such as from a system user, may be captured by a speech input device 102 in a variety of conventional ways. Typically, a microphone or other electro-acoustical device senses speech input from a user and converts it into an analog voltage signal 103 that then is forwarded to a signal processor 104. As is conventionally known, the signal processor 104 includes the necessary analog-to-digital converters, filters, and equalization circuitry and/or software that convert the analog speech input 103 into a digitized stream of data 105 that can be separated into separate units for analysis. Alternatively, this audio data from device 102 can be retrieved from a data storage device. As discussed herein, the system 100 might be realized by suitable hardware and/or software. As such, the blocks shown in FIG. 1B are not meant to indicate separate circuits or to be otherwise limiting, but rather show the functional components of the system.

In particular, the signal processor 104 divides the digital stream of data that is created into a sequence of time-slices, or frames 105, each of which is then processed by a feature generator 106, thereby producing features (vector, matrix, or otherwise organized set of numbers representing the acoustic features of the frames) 107. Further explanation of an example speech recognition system is provided in U.S. Patent No. 4,882,757, entitled "Speech Recognition System", the disclosure of which is incorporated herein by reference in its entirety. This referenced patent discloses Linear Predictive Coding (LPC) coefficients to represent speech; however, other functionally equivalent methods are contemplated within the scope of the invention as well.

A speech recognition search algorithm function 108, realized by an appropriate circuit and/or software in the system 100 analyzes the features 107 in an attempt to determine what hypothesis to assign to the speech input captured by input device 102. As is known in the art in one recognition algorithm, the recognition search 108 relies on probabilistic models provided through 122 from a library of suitable models 110 to recognize the speech input 102. Each of the models in the library 110 may either be customized to a user or be generic to a set of users.

When in operation, the search algorithm 108 (which can be implemented using Hidden Markov Models with a Viterbi algorithm or other modeling techniques such as template matching dynamic time warping (DTW) or neural networks), in essence, compares the features 107 generated in the generator 106 with reference representations of speech, or speech models, in library 110 in order to determine the word or words that best match

the speech input from device 102. Part of this recognition process is to assign a confidence factor for the speech to indicate how closely the sequence of features from the search algorithm 106 matches the closest or best-matching models in library 110. As such, a hypothesis consisting of one or more vocabulary items and associated confidence factors 111 is directed to an acceptance algorithm 112, which also can take as inputs a threshold adjustment 116 and one or more expected responses 114. If the confidence factor is above a predetermined acceptance threshold (or an adjusted threshold when the hypothesis matches the expected response), then the acceptance algorithm 112 makes a decision 118 to accept the hypothesis as recognized speech. If, however, the confidence factor is not above the acceptance threshold, as utilized by the acceptance algorithm, then the acceptance algorithm 112 makes a decision 118 to ignore or reject the recognized speech. The system may then prompt the user to repeat the speech. In this instance, the user may repeat the speech to input device 102. The hypothesis and confidence factors 111, the expected response 114, acceptance algorithm decision 118 and features 107 can also be input to a model adaptation control module 117. Model adaptation control module 117 (which may by implemented in a hardware or software controller or control mechanism) controls the adaptation of library of models 110.

An Example Embodiment Model Adaptation Control Module

FIG. 2 illustrates a model adaptation control module 200, an example embodiment implementation of the model adaptation control module 117 in FIG. 1B. Error rate module 210 (which may be implemented in a processor) determines or estimates an error rate of the speech recognition system. The

error rate may be a word error rate, which can be defined as the percentage or ratio of speech recognition errors over the number of words input to the system, and the error rate can be determined or estimated over a window of time (e.g. predetermined length of time) and/or data (e.g. predetermined number of utterances input to the system). (An utterance is a spoken phrase of at least one word such as '1' or "1-2-3".) Furthermore, the error rate can be an error rate determined or estimated in the following ways: per user; over a number of users; per word; over a set of words; or per a group of consecutively spoken words, such as an utterance, phrase or sentence. Furthermore, the error rate determined by module 210 can be based on actual errors determined from comparing the system's hypothesis to the reference transcript or based on estimated errors deemed to have occurred after evaluating system and user behavior, as discussed later in this application. Therefore, the error rate determination may also be an error rate estimation. Inputs 205 to error rate module 210 are those needed for an error rate determination or estimation used for a particular application. In this example embodiment, inputs 205 are a hypothesis and confidence factor (such as 111 of FIG. 1B) with its associated timing information and expected response(s) (such as 114 of FIG. 1B). Error rate module 210 outputs an error rate 215 to adaptation control module 225.

Adaptation control module 225 controls or adjusts the adaptation of models by model adaptation module 235. Inputs 220 are those needed for the control of model adaptation desired for a particular application. In this example embodiment, inputs 220 are a hypothesis and features (such as 107 of FIG. 1B). Adaptation control module 225 outputs instructions 230 to model

adaptation module 235. In this example embodiment, instructions 230 can include instructions of when to adapt a certain model or models (including instructions to adapt or withhold adaptation), which utterances to use to adapt the models (including the transcription of the utterance and the features observed by the recognition system corresponding to the utterance). By providing control instructions 230, adaptation control module 225, can control when adaptation should occur and determine the criteria to be met before adaptation is ordered. Furthermore, once adaptation is to proceed, the application or adaptation control module may determine whether the existing models are replaced with new models created with the new features (or features of new examples of words) only or whether the existing models are just adapted using information from both the new features and the existing features of the existing models. Model adaptation module 235 outputs adapted models 240 to the library (such as 110 of FIG. 1B) of models.

Because it is useful to prevent the system from adapting in an undesirable way, thereby resulting in a system that performs worse than it did prior to adaptation or a system that degrades over time, and because it is extremely useful to use resources only when necessary, in one embodiment, an adaptation control module (such as 225 of FIG. 2) uses an error rate (such as 215 of FIG. 2) to control or adjust the adaptation (including adapting or withholding adaptation) of models (by for example model adaptation module 235 of FIG. 2).

Example Embodiment Methods for Controlling Model Adaptation

FIG. 3 is a flow chart illustrating a method 300 for controlling or adjusting model adaptation, according to an example embodiment of the

invention. It can be executed by a component of a speech recognition system, such as model adaptation and control module 117 of FIG. 1B. At 305, input speech is received by the speech recognition system from, for example, a user or a recording from a user's session. At 310, initial speech processing is performed (such as processing of the input speech performed by the signal processor 104, feature generator 106 and speech recognition search algorithm 108 of FIG. 1B) for words input to the system. At 315, an error rate, corresponding to either recognition of instances of a word or for recognition of the instances of various words, is determined or estimated. For example, the error rate can be based on recognition errors for the word '1', for the words '1', '2' and '3', for all digits, or for all words in the vocabulary of the system. The error rate can be an updated error rate based on of instances previously and currently input to the system. At 320, a determination is made whether to adapt (such as by the processing of the model adaptation control module 117 of FIG. 1B) a model for the word or various models for the various words, based on the error rate. For example, a determination can be made to adapt the model for the word '1' based on an error rate for the word '1'. In another example, a determination can be made to adapt all words that are digits, based on a combined error rate for all of the digits. If it was determined that the model(s) should not be adapted, next is 305. If the model(s) should be adapted, the model(s) are adapted in 325. After 325 is executed, control returns to 305. Model adaptation in 325 can be performed in the background with control returning to 305 immediately. In other words, the speech recognition system can continue to receive and process speech while the models are being adapted.

In an example embodiment, an error rate is compared to a predetermined criterion such as an error rate threshold to control model adaptation. In other words, an example embodiment makes a comparison of an error rate to an error rate threshold and adapts at least one model or withholds adapting the model based on the comparison. For example, if the error rate 215 is below a particular error rate threshold, an adaptation control module (such as 225 in FIG. 2) provides instructions to model adaptation module 235 to withhold adapting the model(s) associated with the error rate determination or estimation. One reason behind the instructions to adapt is that if the error rate is below a particular error rate threshold, the speech recognition system is performing well and accordingly, model adaptation should not be performed. If the error rate is above a particular error rate threshold, the adaptation control module provides instructions to a model adaptation module to perform model adaptation. The instructions can include performing model adaptation only on models associated with the error rate determination or estimation.

In example embodiments of the invention, the error rate threshold can be a predetermined value, a value settable by a user, a dynamic value, or it can be adjusted upwardly or downwardly. Moreover, the error rate threshold can be based on factors that affect the achievable error rate of the speech recognition system and those that determine an acceptable error rate for the application in which the system is used. Furthermore, the error rate threshold can be based on a number of words in an utterance input to the speech recognition system (or number of words in the system's hypothesis of an utterance), based on environmental factors (such as background noise level

or a signal-to-noise ratio), based on the perplexity of the grammar of a speech recognition system, based on grammar complexity or confusability of the words in the vocabulary, any other measure of difficulty of performing a speech recognition task, or based on a number of words in the vocabulary of a speech recognition system.

Example Embodiments of Error Rates

Throughout this present application, there are various example embodiments for determining or estimating the occurrences of possible (or potential or suspected) errors made by a speech recognition system and an error rate (which can be performed by the error rate module 210 of FIG. 2 and 315 of FIG. 3. The error rate can be used to control or adjust adaptation by the adaptation and control module 225 of FIG. 2 and 320 of FIG. 3). The error rate can be based on any one or combination of the various speech recognition errors discussed in this present application, such as those in the BACKGROUND section of this present application and those discussed below. For example, the error rate can be the ratio of insertion errors over words input to the system. Or for example, the error rate can be the ratio of insertion, substitution and deletion errors over the words input to the system. Or for example, the error rate can be the combination of the low confidence rate and the substitution rates discussed below. The exemplary embodiment error rates discussed below are based on estimated errors which are deemed to have occurred based on evaluating system behavior, the expected response and/or user behavior. Thus, these estimated error rates provide an

advantage of not requiring a reference transcript of words input to the system and comparison of the system's hypotheses corresponding to the words input to the system.

Low Confidence Rate

In an example embodiment of the invention, a count of occurrences of possible errors made by a speech recognition system can be used to determine an estimate of a low confidence rate or an estimate of an error rate. FIG. 4 is a flow chart illustrating a method 400 for identifying errors, which can be executed by components of a speech recognition system, such as the error rate module 210 of FIG. 2. The low confidence rate is the rate at which a word is recognized with a confidence factor within a certain range corresponding to low confidence that the system recognized the word correctly. In other words, the low confidence rate is the frequency at which a word was recognized by the speech recognition system with a confidence factor that is relatively low depending on the application in which the speech recognition system is used. Note that a low confidence rate does not measure errors by the speech recognition system, but the low confidence rate (or a fraction/multiple of its value) can be used in addition to or in place of error rate estimates where error rates (or error rate estimates) are used.

In FIG. 4, at 405, the confidence factor for a hypothesized word is determined. (This confidence factor can be generated by search algorithm 108 of FIG. 1B and supplied to the error rate module 210 of FIG. 2.) At 410, the confidence factor is compared with a range of values corresponding to low confidence that the system recognized the word correctly for the application in which the system is used. If at 410 it is determined that the

confidence factor is outside of the low confidence range, control is returned to 405. If it is determined at 410 that the confidence factor is within the low confidence range, the error count is incremented at 415. After 415, control returns to 405. The error count in 415 may then be combined with counts of other error types to generate an error rate.

An exemplary embodiment, which uses a low confidence rate, also considers when a word is from a hypothesis generated by the system that matches an expected response in counting errors for an error rate estimation. (U.S. Patent Application No. 11/051,825, and the BACKGROUND section of this present application describes scenarios in which an expected response from a user is processed by a speech recognition system. An expected response can be defined as a response that the system expects to receive from the user, as a result of the application in which the system is used). In an example embodiment in the referenced patent application, an acceptance algorithm of the system normally requires that the system's hypothesis is accepted only if a confidence factor for the hypothesis exceeds an acceptance threshold. However, when the system's most likely hypothesis matches an expected response, the hypothesis is more favorably treated so that the hypothesis may be accepted by the system. The reasoning behind the favorable treatment despite the relatively low confidence factor is that a hypothesis matching an expected response usually indicates a high probability of correct recognition.

Turning back to the example embodiment of the present invention, in which the error rate is a low confidence rate, responses that match the expected response and have a relatively low confidence factor for the

application in which the system is used are counted as errors for an error rate estimation. Although a recognition error may not have actually occurred (because the system's hypothesis was correctly accepted due to the hypothesis matching the expected response as described in referenced U.S. Patent Application No. 11/051,825), in this example embodiment, a word with a relatively low confidence is counted as an error for an error rate estimation due to the relatively low confidence factor. The range of confidence factors for which a word is counted as a low confidence could be, for example, between the adjusted acceptance threshold and the original, unadjusted acceptance threshold. More generally, the confidence factor thresholds or range for the counting low confidence errors do not need to match the acceptance threshold and adjusted acceptance threshold in the referenced patent application. The range could be between two other thresholds, including a high confidence threshold, which is higher than the acceptance threshold and indicates the boundary between low and high confidence. In this example embodiment, the range of confidence factors used for the low confidence rate is determined based on the application in which the speech recognition system is used.

Substitution Rate

In an example embodiment of the invention, a count of occurrences of possible substitution errors made by a speech recognition system can be used to determine an estimate of a substitution rate or an estimate of an error rate. The substitution rate is the rate at which substitution errors (such as the substitution errors defined in the BACKGROUND section of this present application) are made by a system. In an exemplary embodiment, a

hypothesis generated by the speech recognition system is compared to an expected response and a substitution error occurs if the system replaces a word in the expected response with an incorrect word in the hypothesis. For example, if the system recognizes "1-5-3" and the expected response is "1-2-3", a substitution error is counted because it is deemed that the system made one substitution: substituting the '5' for the '2'. In other words, if the hypothesis and the expected response do not match word-for-word, but do mostly match (i.e. the hypothesis and the expected response match except for a predetermined number of words), it is a reasonable assumption that a word substitution error has occurred. (The predetermined number of words depends upon the application. For example, an application that uses three-word hypotheses or utterances may define "mostly match" as matching word-for-word except for one word. An application that uses five-word hypotheses or utterances may define "mostly match" as matching word-for-word except for two words.)

Repeated Utterances

Yet in another example embodiment, the error rate is based on a recognition error made by the speech recognition system that is realized after comparing the system's decision on its hypotheses of at least two consecutive or proximate utterances. The decision can occur after the speech recognition system has processed the incoming utterances (such as at 118 of FIG. 1B, after the acceptance algorithm in 112 of FIG. 1B is executed). The recognition error can be for example to reject the system's hypothesis of an incoming utterance, after which the user repeats the utterance, in response to the system's response or lack of one. Or for

example, the recognition error can be to substitute a word that the speech recognition system is unable to recognize correctly, with another word or "garbage" word, in the speech recognition system output. FIGs. 5-6 illustrate example embodiment methods to estimate these types of error rates.

Reject and Repeat

FIG. 5 is a flow chart illustrating a method 500 of an exemplary embodiment for identifying possible occurrences of errors made by a speech recognition system. The count of the possible occurrences of errors can be used to determine an estimate of an error rate. Method 500 can be executed by a component of a speech recognition system, such as error rate module 210 of FIG. 2. In this embodiment, the determination of whether the speech recognition system made an error is made when the speech recognition system receives at least two consecutive or proximate utterances. The system and user behavior is as follows: the system rejects its hypothesis of the first utterance; the user repeats the first utterance in the second utterance; and the system accepts its hypothesis of the second utterance. The first and second hypotheses generated by the system substantially match. In other words, the hypotheses match word-for-word but a hypothesis may also include a recognized model that is considered to be negligible for this particular error determination. For example, a hypothesis could include a recognized model indicating a user's breath or sigh and these recognized models can be considered negligible for this particular error determination. However, recognized models in a hypothesis that indicate the system is having difficulty discerning what the user spoke (such as for example a recognized model indicating silence, garbage or a wild card word) might not

be considered negligible. (The determination of whether a recognized model is negligible depends upon the particular speech recognition system and the application in which it is used.) An example is as follows: a user speaks a first utterance "1-2-3"; the system correctly recognizes it (i.e. generates a hypothesis of "1-2-3") but rejects its hypothesis because of a low confidence factor; the user repeats "1-2-3" in a second utterance and the system correctly recognizes it (i.e. generates a hypothesis of "1-2-3") and accepts its hypothesis. A rationale behind this type of error detection mechanism is that if the two matching utterances are spoken consecutively, and the system accepts its hypothesis of the second utterance, one could reasonably assume that the system should have accepted its hypothesis of the first utterance and that it erred in not doing so. This heuristic can alternatively require that the two utterances are spoken within a predetermined amount of time of each other, or further refined by also requiring that the utterances are spoken consecutively.

In FIG. 5, at 505, decisions made by a speech recognition system (such as at 118 of FIG. 1B) on a first and second utterance are received for processing by a model adaptation and control module (such as 117 of FIG. 1B). At 510, verifications are performed. These verifications can include one or more of the following conditions: verifying that the system's hypotheses of those utterances contain multiple words; verifying that the system's hypothesis of the second utterance contains all accepted words; verifying that there was at least one rejected word in the system's hypothesis for the first utterance; and verifying that the second hypothesis matches the expected response (if there is one). At 515, if the verifications pass, next is 520.

Otherwise, control returns to 505. At 520, the words in the first and second hypotheses are compared word-for-word to find if they match. For example, if the first hypothesis is "one-two-three" and the second hypothesis is "one-three-three", there is a mismatch. If the hypotheses match word-for-word, there is a high probability that an incorrect rejection error has occurred, with the reasoning that the user repeated himself and the system recognized the second utterance correctly. If the hypotheses match word-for-word, next is 525. Otherwise, control returns to 505. At 525, the error count is incremented and control returns to 505. The error count in 525 may then be combined with counts of other error types to generate an overall error rate.

Substitute and Repeat

FIG. 6 is a flow chart illustrating a method 600 of an exemplary embodiment for identifying possible occurrences of errors made by a speech recognition system. The count of the possible occurrences of errors can be used to determine an estimate of an error rate. Method 600 can be executed by a component of a speech recognition system, such as error rate module 210 of FIG. 2. In this embodiment, the determination of whether the speech recognition system made an error is made when the speech recognition system receives at least two consecutive or proximate utterances and the system substitutes a word in its hypothesis of the first utterance and recognizes and accepts all of the words in its hypothesis of the second utterance. An example is as follows: a user speaks a first utterance "1-2-3"; the system misrecognizes it (e.g. generates a hypothesis "1-5-3") and accepts its hypothesis; the user repeats "1-2-3" in a second utterance within a proximity of the first utterance; the system correctly recognizes it (i.e.

generates a hypothesis "1-2-3") and accepts its hypothesis. A rationale behind this method of detecting errors is that if the two utterances are spoken consecutively or within a proximity of each other, and if the system accepts its hypothesis of the second utterance, then the system likely made a substitution in its hypothesis of the first utterance. There are heuristics that may be used to guard against the system considering consecutive or proximate recognitions differing by a single word as containing a substitution error when in fact they do not. The heuristics include checking for one or more of the following possible conditions: there were no intervening utterances that indicate that the first utterance was correctly recognized by the system; the two utterances being compared represent the same piece of information being entered into the system, for example, the two utterances being compared occurred at the same position in the dialogue between the user and the recognition system, or in response to the same prompt; the two utterances were spoken within a predetermined amount of time or in other words the time between the two utterances being compared was short enough to indicate that the user was repeating the initial utterance.

In FIG. 6, at 605, decisions made by a speech recognition system (such as at 118 of FIG. 1B) on a first and second utterance are received for processing by a model adaptation and control module (such as 117 of FIG. 1B). At 610, verifications are performed. These verifications improve the accuracy of the estimate of the substitution error rate and may include: verifying that the utterances were spoken consecutively or within a proximity of each other; verifying that the system's hypotheses of the utterances contain multiple words; verifying that the system's hypotheses of the

utterances contain all accepted words; verifying that the user was prompted for the same information by the system both times; verifying that the first hypothesis does not match the expected response (if there is one); and verifying that the second hypothesis does match the expected response (if there is one); and checking for a condition indicating a substitution error occurred (such as those described above). At 615, the words in the system's hypotheses of the first and second utterances are compared word-for-word to see if they match. If the hypotheses do not match word-for-word, next is 620. Otherwise, control returns to 605. At 620, if the verifications pass, next is 625. Otherwise, control returns to 605. At 625, the words in the system's hypotheses of the first and second utterances are compared word-for-word to find how closely they match. For example, if the first hypothesis is "1-2-3" and the second hypothesis is "1-5-3", there is a mismatch of one word. In this case, the '5' was substituted for the '2'. If the hypotheses do not match word-for-word, but do mostly match, (e.g. the hypotheses match except for one word), it is a reasonable assumption that a word substitution error has occurred, with the reasoning that the system performed verifications such as checking for at least one condition indicating a substitution error occurred, the user repeated the same utterance, the system recognized the second utterance correctly, and the system incorrectly substituted a word in its hypothesis of the first utterance. (The definition of "mostly match" depends upon the application. For example, an application that uses five-word hypotheses or utterances may define "mostly match" as matching word-for-word except for two words.) If the hypotheses mostly match word-for-word,

next is 630 where the error count is incremented followed by control returning to 605. The error count in 630 may then be combined with counts of other error types to generate an overall error rate.

The same approach as in the previous paragraph can be used to detect deletion due to garbage errors where a content word is recognized by the system as garbage in a first utterance, then correctly recognized in the next utterance. By comparing the recognition results of the two utterances and using verifications such as those described above, one can detect the error. For example, if the system's hypothesis of the first utterance is "1-GARBAGE-3" and the system's hypothesis of the second utterance is "1-5-3", there is a mismatch of one word, and it becomes a reasonable assumption that the speech recognition system made an error in its hypothesis of the first utterance. Again, similar verifications as described above may be used to guard against the system considering a correct recognition to be in error.

The same approach as described above in the discussion of FIG. 6 can also be used to detect other types of errors, such as a deletion due to rejected substitution error. An example of a deletion due to rejected substitution error is as follows. A user speaks a first utterance "1-2-3" and the system recognizes it (i.e. generates a hypothesis "1-2-3"), but the system rejects the '2' in its hypothesis. The user speaks a second utterance "1-5-3" within a proximity of the first utterance and the system recognizes it (i.e. generates a hypothesis "1-5-3") and the system accepts its hypothesis. When the system rejected the '2' in its hypothesis of the first utterance, the system made a deletion due to rejected substitution error. The method for detecting this type of error is similar to that described in the discussion of FIG.

6, with the difference that the system's hypothesis of the first utterance does not need to contain all accepted words.

Correction Rate

In an exemplary embodiment error rate for a speech recognition system, a count of occurrences of when a user provides feedback to the system can be used as an estimate of an error rate or an estimate for part of an error rate. The reasoning behind using a such a count to estimate of an error rate or estimate part of an error rate is that when a correction is commanded to the system, it may indicate that an error occurred. Examples of user feedback are described in the BACKGROUND section of this present application. The count can include the rate at which the user indicates that the system made a mistake. Furthermore, the user may provide feedback in response to the system requesting feedback, such as asking the user to confirm a hypothesis generated by the system or asking the user to identify what word was spoken by the user. The feedback may include a word indicating aggravation by the user or the feed back may be a correction command to the system, such as "back-up" or "erase". In determining or estimating the error rate, considerations can be made for the amount of time and data needed to determine or estimate an error rate that is useful for the application in which the speech recognition system is used. One example consideration is that the error rate is determined or estimated for speech input to the speech recognition system over a predetermined length of time. Another example consideration is that the error rate is determined or estimated for speech input to the speech recognition system over a predetermined number of utterances, words, or hypotheses.

Error Rate Considerations

Another example consideration is that the error rate is determined or estimated from hypotheses of utterances collected over a moving or sliding window or a collection period that is dynamic in length of time and/or size of data. As a result, the error rate is determined or estimated over a period when useful data has been collected. For example, a moving or sliding window can cover a collection of data taken from equal periods in noisy environment and a quiet environment to offset any favoring by the speech recognition system in one of those environments. Other examples of moving, sliding windows are those that collect data only during recent use (e.g. the last half-hour) of the speech recognition system, collecting data for time spent by a particular user (e.g. since the user started using the system), or collecting a certain amount of data regardless of the time spent collecting the data (e.g. the last one-hundred hypotheses).

It can be understood by those skilled in the art that in other exemplary embodiments of the invention, other error rates can be used, such as a phrase error rate, utterance error rate, and sentence error rate. For example, an utterance error rate can be defined as the percentage or ratio of speech recognition errors over the number of utterances input to the system and the utterance error rate can be used in controlling or adjusting an adaptation of a model.

Furthermore, it will be evident to one skilled in the art that the various methods to identify possible errors can process the same hypotheses, but safeguards must be taken to avoid double counting, as some possible errors may be counted by more than one method.

Example Embodiment of Model Adaptation

FIG. 7 is a flow chart illustrating a method 700 for model adaptation, according to an example embodiment of the invention. It can be executed by a component of a speech recognition system, such as the model adaptation module 235 of FIG. 2. In an example embodiment, method 700 is executed under the control of a control module (such as 220 of FIG. 2). In other words, method 700 is performed based on instructions from a control module. The instructions can include instructions of when to adapt a certain model or multiple models (including instructions to adapt or withhold adaptation) and information corresponding to the model(s) to be adapted, such as transcription of an utterance and a set of features observed by the speech recognition system corresponding to the utterance.

At 705, the features observed by a speech recognition system corresponding to an input utterance are aligned with the states in the models for the words of the utterance. In an example embodiment, the Baum-Welch re-estimation algorithm can be used to perform the alignment. At 710, the statistics (for example, means and variances) of the states are updated using the values of the features. At 715, these values are mixed into the models with an appropriate weighting to maintain a balance between previous training data and new features. Thus, in an example embodiment of the invention, new models are created through adaptation by using the observed features of an input utterance to adapt existing or original models. In that scenario, both the observed features of the input utterance and the existing features of the original models, and the statistics associated with each, are used to create the new models. Additionally, in such a scenario, the new statistics might be

weighted in various fashions to tailor their effect on the original statistics in the model. In an alternative example embodiment of the invention, only the new observed features (of new examples of words), and information therefrom, are utilized to create the new model. That is, a new model is created with the new features only. Furthermore, the adaptation could be performed using data from a single user or multiple users. In one particular embodiment, only speech data from an individual user might be used to perform the adaptation. This generates a model that is adapted for that user and performs well for that user.

FIGs. 3-7 show flow charts illustrating methods according to embodiments of the invention. The techniques illustrated in these figures may be performed sequentially, in parallel or in an order other than that which is described. It should be appreciated that not all of the techniques described are required to be performed, that additional techniques may be added, and that some of the illustrated techniques may be substituted with other techniques.

The invention in its various forms may be implemented directly in the software of a speech recognition system. That is, the improvements are actually part of the speech recognition system. Alternatively, the invention does not have to be built into the speech recognition system. Rather, the invention or parts of the invention may be implemented in a separate module, program or application which interacts with a speech recognition system to provide the benefits of the invention. For example, a separate application or software module may be utilized to handle the adaptation in accordance with

the principles of the invention. Specifically, an application may interface with a speech recognition system to determine or estimate an error rate and/or control when and how models are adapted.

In the foregoing description, the invention is described with reference to specific example embodiments thereof. The specification and drawings are accordingly to be regarded in an illustrative rather than in a restrictive sense and it is not the intention of the applicants to restrict or in any way limit the scope of the appended claims to such detail. It will, however, be evident to those skilled in the art that additional advantages and modifications can be made, in a computer program product or software, hardware or any combination thereof, without departing from the broader spirit and scope of the invention will readily appear. Software embodiments may include an article of manufacture on a machine accessible or machine readable medium having instructions. Furthermore, software embodiments may be distributed or downloaded via a network or email. The instructions on the machine accessible or machine readable medium may be used to program a computer system, such as for example, a PC, cell phone, industrial mobile computer, PDA, electronic headset or other electronic device with exemplary embodiment methods or approaches disclosed herein. The machine-readable medium may include, but is not limited to non volatile memory, floppy diskettes, optical disks, CD-ROMs, and magneto-optical disks or other type of media/machine-readable medium suitable for storing or transmitting electronic instructions. Furthermore, departures may be made from the application in which the invention is described without departing from the spirit and scope of the invention. For example, the example speech recognition

system described herein has focused on wearable terminals. However, the principles of the invention are applicable to other speech recognition environments as well.

CLAIMS

What is claimed is:

1. A method for model adaptation for a speech recognition system comprising:

determining an error rate, corresponding to either recognition of instances of a word or recognition of instances of various words among a set of words; and

adjusting an adaptation of a model for the word or various models for the various words, based on the error rate.

2. The method of Claim 1, wherein the error rate is based on actual errors determined from comparing a transcript of words input to the system with corresponding hypotheses generated by the system.

3. The method of Claim 1, wherein the error rate is based on estimated errors estimated from evaluating system and user behavior corresponding to proximate utterances input to the system by the user.

4. The method of Claim 1, wherein adjusting an adaptation comprises adapting the model or the various models or withholding adapting the model or the various models based on the error rate.

5. The method of Claim 1, wherein adjusting the adaptation comprises:

making a comparison of the error rate to an error rate threshold; and

adapting the model or the various models or withholding adapting the model or the various models based on the comparison.

6. The method of Claim 1, wherein adjusting the adaptation comprises:

withholding adapting the model or the various models until the error rate meets an error rate threshold.

7. The method of Claim 5, wherein the error rate threshold is a predetermined value.

8. The method of Claim 5, wherein the error rate threshold is a value settable by a user.

9. The method of Claim 5, wherein the error rate threshold is a dynamic value.

10. The method of Claim 5, wherein the error rate threshold is based on a number of words in a hypothesis of an utterance input to the system.

11. The method of Claim 5, wherein the error rate threshold is based on an environmental factor.

12. The method of Claim 5, wherein the error rate threshold is based on a measure of difficulty of the recognition of the word.

13. A method for identifying a possible error made by a speech recognition system, comprising:

identifying an instance of a word that was recognized by the system within a certain confidence factor range.

14. The method of Claim 13, wherein a count of occurrences of the possible error is used in adjusting an adaptation a model for the word associated with the possible error.

15. The method of Claim 13, wherein the word is from a hypothesis generated by the system that matches an expected response.

16. The method of Claim 13, wherein the range comprises values corresponding to low confidence that the system recognized the instances of the word correctly for an application in which the system is used.

17. The method of Claim 13, wherein the range is a predetermined set of values in between a low confidence threshold and a high confidence threshold.

18. The method of Claim 17, wherein the low confidence threshold is an adjusted acceptance threshold, which is a value lower than the acceptance threshold.

19. The method of Claim 17, wherein the high confidence threshold is a value equal to or higher than an acceptance threshold for accepting a hypothesis generated by the system, in which a confidence value associated with the hypothesis must exceed the acceptance threshold in order for the hypothesis to be accepted.

20. A method for identifying a possible error made by a speech recognition system comprising:

identifying an instance where the system rejects a first hypothesis of a first utterance, followed by the system accepting a second hypothesis of a second utterance, wherein the first and second hypotheses substantially match word-for-word.

21. The method of Claim 20, wherein the first and second hypotheses substantially match word-for-word by matching word-for-word, except that one of the first and second hypotheses includes at least one additional recognized model that is negligible for purposes of identifying the possible error.

22. The method of Claim 20, wherein a count of occurrences of the possible error is used in adjusting an adaptation of a model for a word associated with the possible error.

23. The method of Claim 20, wherein the system rejects the first hypothesis due to a confidence factor of the first hypothesis not exceeding an acceptance threshold.

24. The method of Claim 20, wherein the system rejects at least one word in the first hypothesis.

25. The method of Claim 20, wherein the first and second utterances are spoken consecutively, proximately or within a predetermined amount of time.

26. A method for identifying a possible error made by a speech recognition system comprising:

identifying when the system generates first and second hypotheses of two utterances and the system accepts the second hypothesis, wherein the two hypotheses do not match word-for-word, but the hypotheses mostly match word-for-word.

27. The method of Claim 26, wherein the hypotheses mostly match word-for-word by matching word-for-word except for a predetermined number of words.

28. The method of Claim 26, wherein a count of occurrences of the possible error is used in adjusting an adaptation a model for a word associated with the possible error.

29. The method of Claim 26, wherein the two utterances are spoken consecutively, proximately, or within a predetermined amount of time.

30. The method of Claim 26, wherein there are no intervening utterances that indicate that the first of the two utterances was correctly recognized by the system.

31. The method of Claim 26, wherein the two hypotheses differ in that a word in the second hypothesis is substituted with a word in the first hypothesis.

32. The method of Claim 26, wherein the hypotheses differ in that a word in the second hypothesis is substituted with garbage in the first hypothesis.

33. A method for identifying a possible error made by a speech recognition system, comprising:

identifying when a hypothesis generated by the system does not match an expected response word-for-word, but the hypothesis mostly matches the expected response word-for-word.

34. The method of Claim 33, wherein a count of occurrences of the possible errors is used in adjusting adaptation of a model for a word associated with the possible errors.

35. The method of Claim 33, wherein the hypothesis mostly matches the expected response word-for-word by matching word-for-word except for one word.

36. A method for adapting a model for a speech recognition system comprising:

generating a count of occurrences of when a user provides feedback to the system; and

adjusting an adaptation of the model based on the count.

37. The method of Claim 36, wherein the count is determined without using a transcript of words input to the system.

38. The method of Claim 36, wherein the count of occurrences includes a count of occurrences of when the user indicates that the system made a mistake.

39. The method of Claim 36, wherein the user provides the feedback in response to the system requesting the feedback.

40. The method of Claim 39, wherein the feedback includes a word or other indication of aggravation by the user.

41. The method of Claim 39, wherein the feedback is a correction command to the system.

42. An apparatus for model adaptation for a speech recognition system comprising:

a processor adapted to determine an error rate, corresponding to either recognition of instances of a word or recognition of instances of various words among a set of words; and

a controller adapted to adjust an adaptation of a model for the word or various models for the various words, based on the error rate.

43. The apparatus of Claim 42, wherein the error rate is based on actual errors determined from comparing a transcript of words input to the system with corresponding hypotheses generated by the system.

44. The apparatus of Claim 42, wherein the error rate is based on estimated errors estimated from evaluating system and user behavior corresponding to proximate utterances input to the system by the user.

45. The apparatus of Claim 42, wherein the controller adjusts an adaptation by adapting the model or the various models or withholding adapting the model or the various models based on the error rate.

46. The apparatus of Claim 42, wherein the controller adjusts the adaptation by

making a comparison of the error rate to an error rate threshold; and

adapting the model or the various models or withholding adapting the model or the various models based on the comparison.

47. The apparatus of Claim 42, wherein the controller adjusts the adaptation by withholding adapting the model or the various models until the error rate meets an error rate threshold.

48. The apparatus of Claim 42, wherein the error rate threshold is a predetermined value.

49. The apparatus of Claim 42, wherein the error rate threshold is a value settable by a user.

50. The apparatus of Claim 42, wherein the error rate threshold is a dynamic value.

51. The apparatus of Claim 42, wherein the error rate threshold is based on a number of words in a hypothesis of an utterance input to the system.

52. The apparatus of Claim 42, wherein the error rate threshold is based on an environmental factor.

53. The apparatus of Claim 42, wherein the error rate threshold is based a measure of difficulty of the recognition of the word.

54. An apparatus for identifying a possible error made by a speech recognition system, comprising:

a processor adapted to identify an instance of a word that was recognized by the system within a certain confidence factor range.

55. The apparatus of Claim 54, wherein a count of occurrences of the possible error is used in adjusting adaptation of a model for the word associated with the possible error.

56. The apparatus of Claim 54, wherein the word is from a hypothesis generated by the system that matches an expected response.

57. The apparatus of Claim 54, wherein the range comprises values corresponding to low confidence that the system recognized the instances of the word correctly for an application in which the system is used.

58. The apparatus of Claim 54, wherein the range is a predetermined set of values in between a low confidence threshold and a high confidence threshold.

59. The apparatus of Claim 58, wherein the high confidence threshold is an acceptance threshold for accepting a hypothesis generated by the system, in which a confidence threshold associated with the hypothesis must exceed the acceptance threshold in order for the hypothesis to be accepted.

60. The apparatus of Claim 58, wherein the low confidence threshold is an adjusted acceptance threshold, which is a value lower than the acceptance threshold.

61. The apparatus of Claim 58, wherein the high confidence threshold is a value equal to or higher than an acceptance threshold for accepting a hypothesis generated by the system, in which a confidence value associated with the hypothesis must exceed the acceptance threshold in order for the hypothesis to be accepted.

62. An apparatus for identifying a possible error made by a speech recognition system comprising:

a processor adapted to identify an instance where the system rejects a first hypothesis of a first utterance, followed by the system accepting a second hypothesis of a second utterance, wherein the first and second hypotheses substantially match word-for-word.

63. The apparatus of Claim 62, wherein the first and second hypotheses substantially match word-for-word by matching word-for-word, except that one of the first and second hypotheses includes at least one additional recognized model that is negligible for purposes of identifying the possible error.

64. The apparatus of Claim 62, wherein a count of occurrences of the possible error is used in adjusting an adaptation of a model for a word associated with the possible error.

65. The apparatus of Claim 62, wherein the system rejects the first hypothesis due to a confidence factor of the first hypothesis not exceeding an acceptance threshold.

66. The apparatus of Claim 62, wherein the system rejects at least one word in the first hypothesis.

67. The apparatus of Claim 62, wherein the first and second utterances are spoken consecutively, proximately or within a predetermined amount of time.

68. An apparatus for identifying a possible error made by a speech recognition system comprising:

a processor adapted to identify when the system generates first and second hypotheses of two utterances and the system accepts the second hypothesis, wherein the two hypotheses do not match word-for-word, but the hypotheses mostly match word-for-word.

69. The apparatus of Claim 68, wherein the hypotheses mostly match word-for-word by matching word-for-word except for a predetermined number of words.

70. The apparatus of Claim 68, wherein a count of occurrences of the possible error is used in adjusting an adaptation a model for a word associated with the possible error.

71. The apparatus of Claim 68, wherein the two utterances are spoken consecutively, proximately, or within a predetermined amount of time.

72. The apparatus of Claim 68, wherein there are no intervening utterances that indicate that the first of the two utterances was correctly recognized by the system.

73. The apparatus of Claim 68, wherein the two hypotheses differ in that a word in the second hypothesis is substituted with a word in the first hypothesis.

74. The apparatus of Claim 68, wherein the hypothesis differ in that a word in the second hypothesis is substituted with garbage in the first hypothesis.

75. An apparatus for identifying a possible error rate made by a speech recognition system, comprising:

a processor adapted to identify when a hypothesis generated by the system does not match an expected response word-for-word, but the hypothesis mostly matches the expected response word-for-word.

76. The apparatus of Claim 75, wherein a count of occurrences of the possible error is used in adjusting an adaptation of a model for a word associated with the possible error.

77. The apparatus of Claim 75, wherein the hypothesis mostly matches the expected response word-for-word by matching word-for-word except for a predetermined number of words.

78. An apparatus for adapting a model for a speech recognition system comprising:

a processor adapted to generate a count of occurrences of when a user provides feedback to the system; and

a controller adapted to adjust an adaptation of the model based on the count.

79. The apparatus of Claim 78, wherein the count is determined without using a transcript of words input to the system.

80. The apparatus of Claim 78, wherein the transcript is compared with corresponding hypotheses generated by the system.

81. The apparatus of Claim 78, wherein the count includes a count of occurrences of when the user indicates that the system made a mistake.

82. The apparatus of Claim 78, wherein the user provides the feedback in response to the system requesting the feedback.

83. The apparatus of Claim 82, wherein the feedback includes a word indicating aggravation by the user.

84. The apparatus of Claim 82, wherein the feedback is a correction command to the system.

85. A method for estimating an error rate for a speech recognition system comprising:

estimating a rate at which the speech recognition system rejects a first utterance, followed by a user repeating the first utterance in a second utterance, followed by the system accepting the second utterance.

86. The method of Claim 85, wherein the rate is used in a determination of whether to adapt a model associated with the error rate.

87. The method of Claim 85, wherein the system rejects the first utterance due to a confidence factor of the system's hypothesis for the first utterance not exceeding an acceptance threshold.

88. The method of Claim 85, wherein the system rejects at least one word in the first utterance.

89. The method of Claim 85, wherein the first and second utterances are spoken within a predetermined amount of time.

90. The method of Claim 85, wherein the first and second utterances are spoken consecutively or proximately.

91. A method for estimating an error rate for a speech recognition system comprising:

estimating a rate at which the system substitutes a word in the system's hypothesis for a first utterance, followed by a user repeating the first utterance in a second utterance, followed by the system recognizing and accepting the second utterance.

92. The method of Claim 91, wherein the rate is used in a determination of whether to adapt a model associated with the word.

93. The method of Claim 91, wherein the first and second utterances are spoken consecutively or proximately.

94. The method of Claim 91, wherein there are no intervening utterances that indicate that the first utterance was correctly recognized by the system.

95. The method of Claim 91, wherein the first and second utterances represent a same piece of information being entered into the system.

96. The method of Claim 91, wherein the first and second utterances were spoken within a predetermined amount of time.

97. The method of Claim 91, wherein the system substitutes the word in the first utterance with another word.

98. The method of Claim 91, wherein the system substitutes the word in the first utterance with garbage.

99. A method for estimating an error rate for a speech recognition system, comprising:

estimating a substitution rate at which the system makes a substitution error, when comparing a hypothesis generated by the system to an expected response.

100. The method of Claim 99, wherein the substitution rate is used in a determination of whether to adapt a model associated with the substitution error.

101. The method of Claim 99, wherein the substitution error is a replacement of one word that is in the expected response with an incorrect word in the hypothesis.

102. A apparatus for estimating an error rate for a speech recognition system comprising:

a processor adapted to estimate a rate at which the speech recognition system rejects a first utterance, followed by a user repeating the first utterance in a second utterance, followed by the system accepting the second utterance.

103. The apparatus of Claim 102, wherein the rate is used in a determination of whether to adapt a model associated with the error rate.

104. The apparatus of Claim 102, wherein the system rejects the first utterance due to a confidence factor of the system's hypothesis for the first utterance not exceeding an acceptance threshold.

105. The apparatus of Claim 102, wherein the system rejects at least one word in the first utterance.

106. The apparatus of Claim 102, wherein the first and second utterances are spoken within a predetermined amount of time.

107. The apparatus of Claim 102, wherein the first and second utterances are spoken consecutively or proximately.

108. An apparatus for estimating an error rate for a speech recognition system comprising:

a processor adapted to estimate a rate at which the system substitutes a word in the system's hypothesis for a first utterance, followed by a user repeating the first utterance in a second utterance, followed by the system recognizing and accepting the second utterance.

109. The apparatus of Claim 108, wherein the rate is used in a determination of whether to adapt a model associated with the word.

110. The apparatus of Claim 108, wherein the first and second utterances are spoken consecutively or proximately.

111. The apparatus of Claim 108, wherein there are no intervening utterances that indicate that the first utterance was correctly recognized by the system.

112. The apparatus of Claim 108, wherein the first and second utterances represent a same piece of information being entered into the system.

113. The apparatus of Claim 108, wherein the first and second utterances were spoken within a predetermined amount of time.

114. The apparatus of Claim 108, wherein the system substitutes the word in the first utterance with another word.

115. The apparatus of Claim 108, wherein the system substitutes the word in the first utterance with garbage.

116. An apparatus for estimating an error rate for a speech recognition system, comprising:

a processor adapted to estimate a substitution rate at which the system makes a substitution error, when comparing a hypothesis generated by the system to an expected response.

117. The apparatus of Claim 116, wherein the substitution rate is used in a determination of whether to adapt a model associated with the substitution error.

118. The apparatus of Claim 116, wherein the substitution error is a replacement of one word that is in the expected response with an incorrect word in the hypothesis.

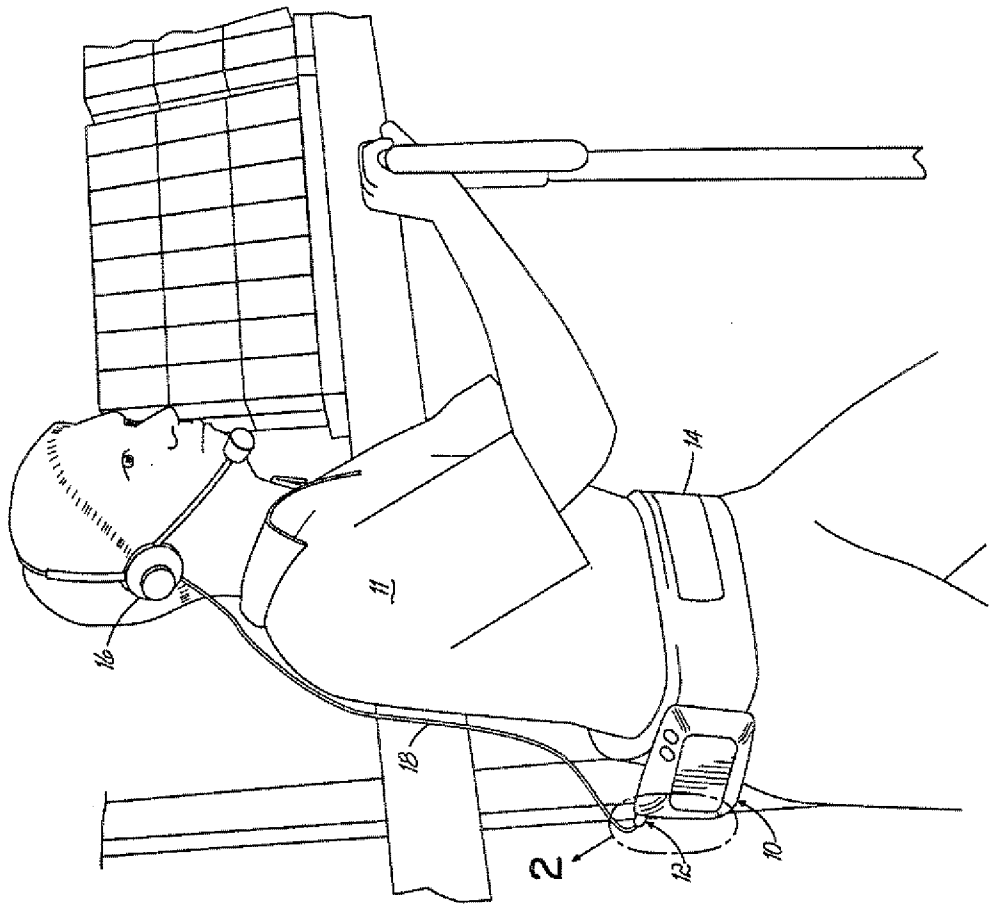


FIG. 1A

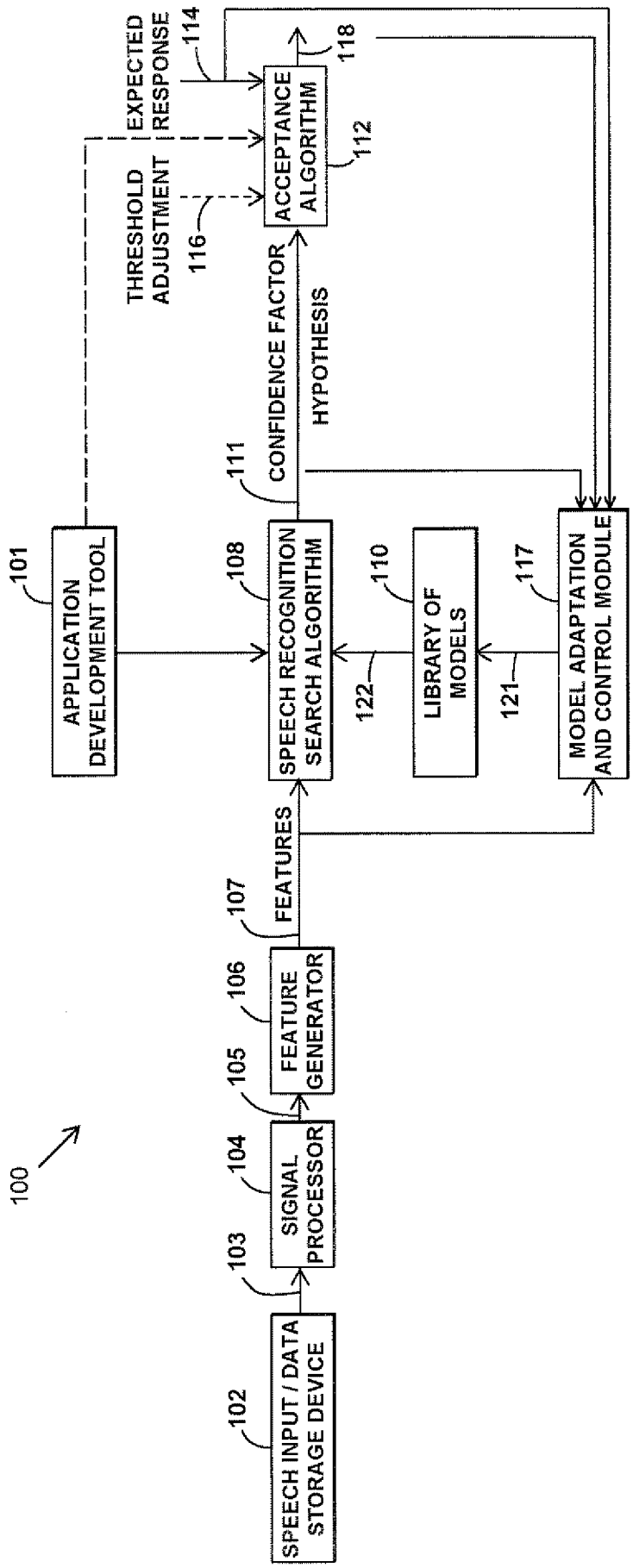

100
SPEECH INPUT / DATA STORAGE DEVICE
102
SIGNAL PROCESSOR
103
104
FEATURE GENERATOR
105
106
FEATURES
107
APPLICATION DEVELOPMENT TOOL
101
SPEECH RECOGNITION SEARCH ALGORITHM
108
LIBRARY OF MODELS
110
MODEL ADAPTATION AND CONTROL MODULE
117
122
121
CONFIDENCE FACTOR
111
HYPOTHESIS
THRESHOLD ADJUSTMENT
116
EXPECTED RESPONSE
114
ACCEPTANCE ALGORITHM
112
118

FIG. 1B

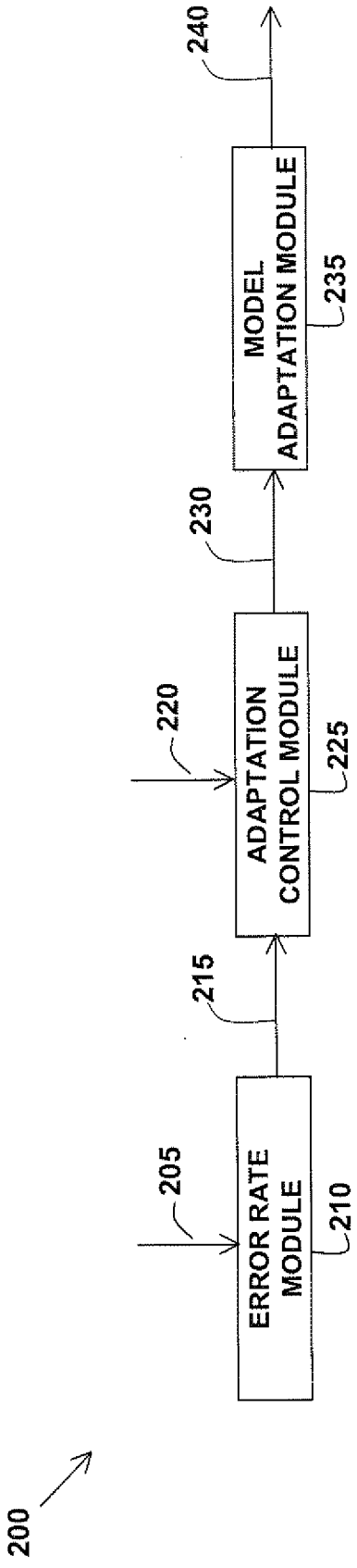

200
205
ERROR RATE MODULE
210
215
ADAPTATION CONTROL MODULE
220
225
230
MODEL ADAPTATION MODULE
235
240

FIG. 2

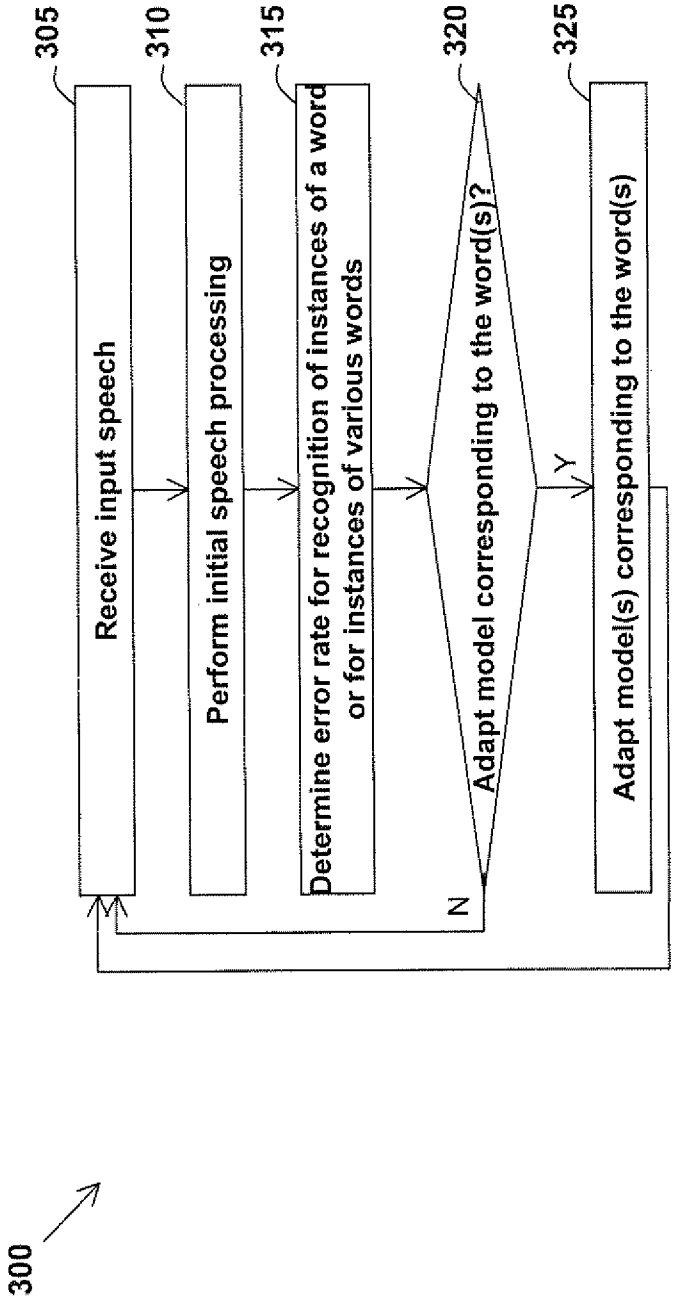

300
305
Receive input speech
310
Perform initial speech processing
315
Determine error rate for recognition of instances of a word or for instances of various words
320
Adapt model corresponding to the word(s)?
Y
N
325
Adapt model(s) corresponding to the word(s)

FIG. 3

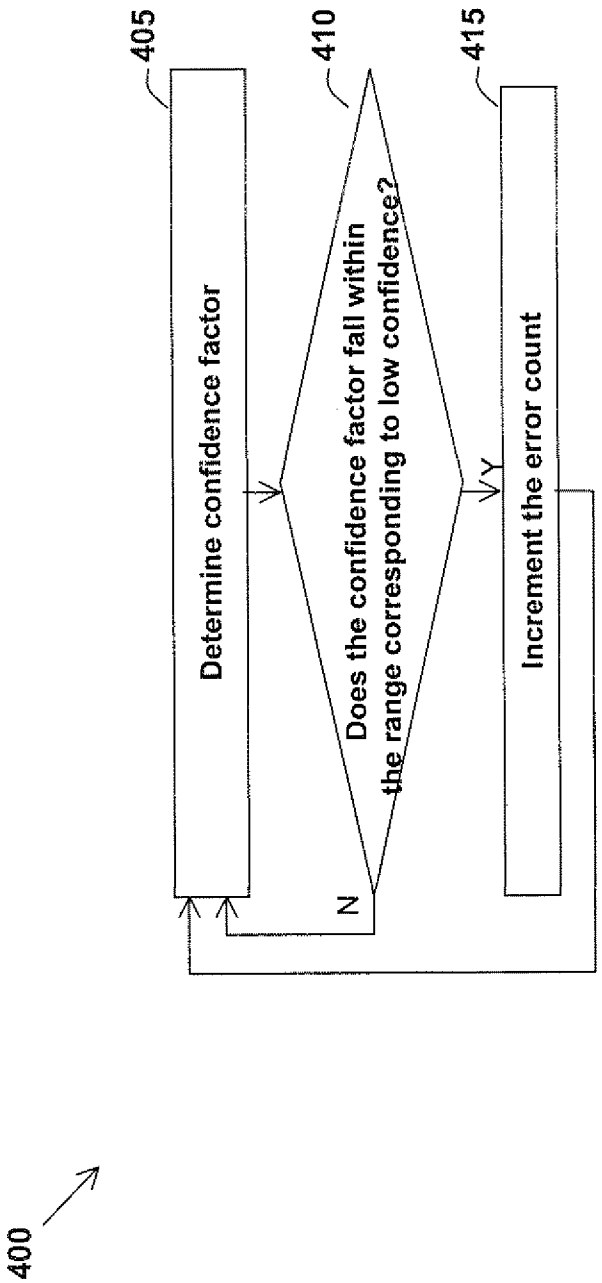

400
405
Determine confidence factor
410
Does the confidence factor fall within the range corresponding to low confidence?
N
Y
415
Increment the error count

FIG. 4

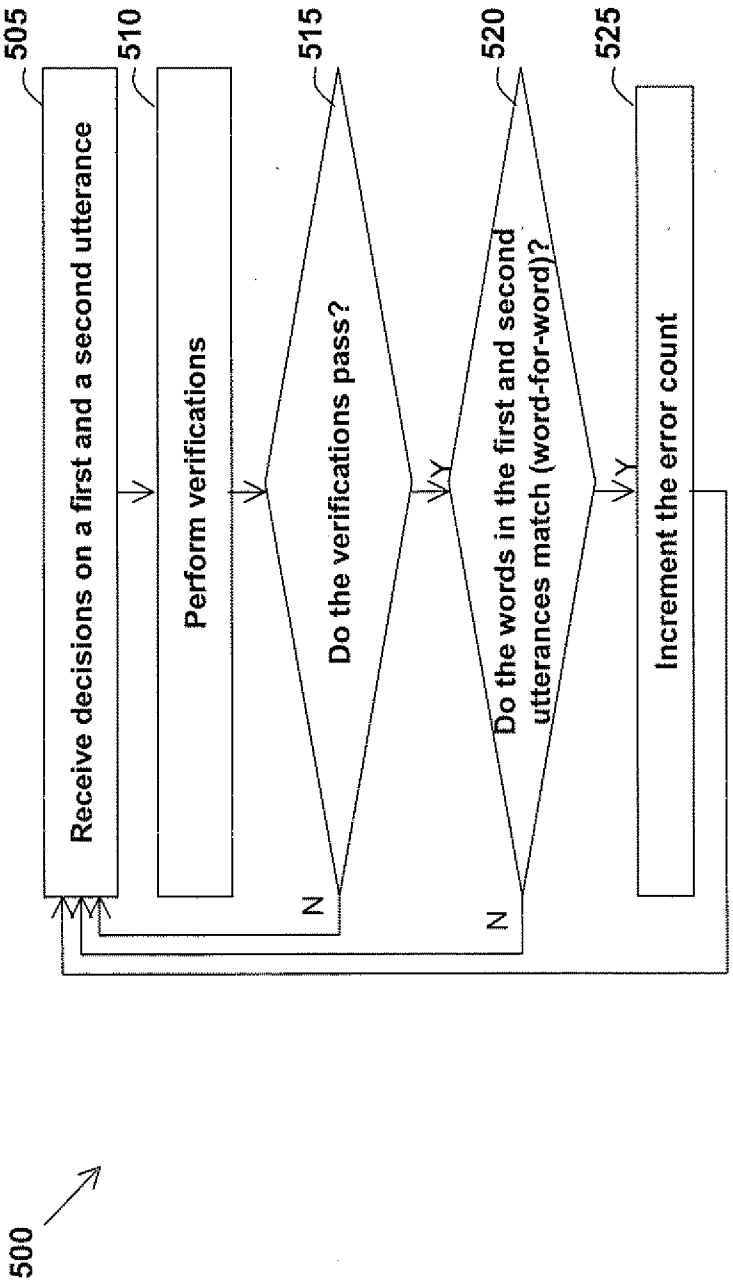

505
Receive decisions on a first and a second utterance
510
Perform verifications
515
Do the verifications pass?
N
520
Do the words in the first and second utterances match (word-for-word)?
N
525
Increment the error count
500

FIG. 5

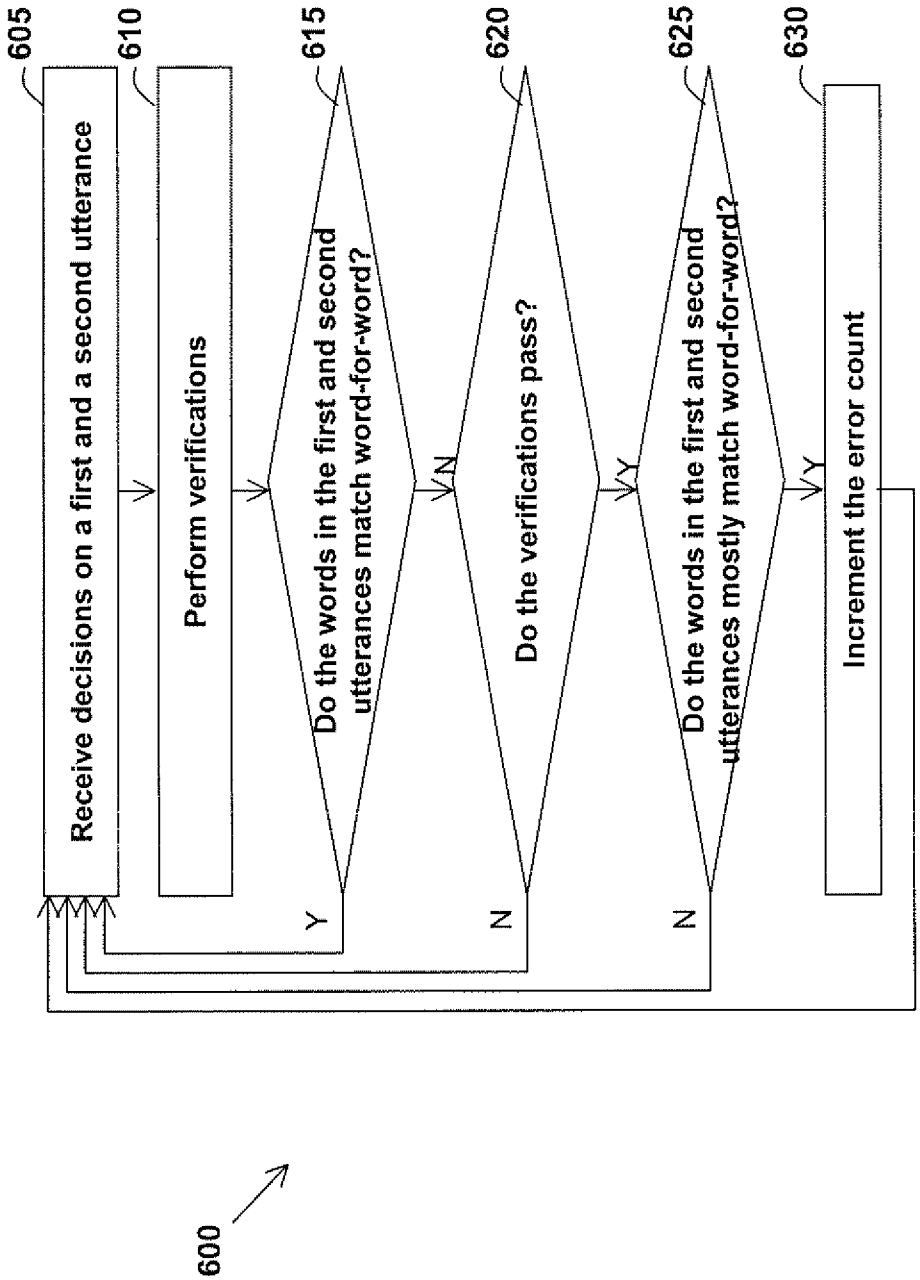

605
Receive decisions on a first and a second utterance
610
Perform verifications
615
Do the words in the first and second utterances match word-for-word?
Y
620
Do the verifications pass?
N
625
Do the words in the first and second utterances mostly match word-for-word?
N
Y
630
Increment the error count
Y
600

FIG. 6

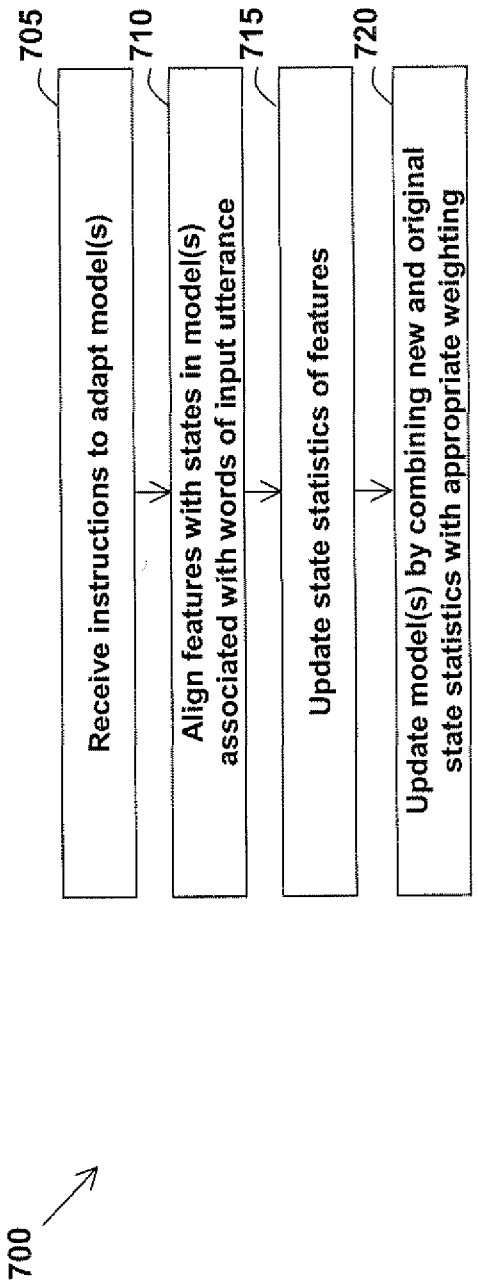

700
705
Receive instructions to adapt model(s)
710
Align features with states in model(s) associated with words of input utterance
715
Update state statistics of features
720
Update model(s) by combining new and original state statistics with appropriate weighting

FIG. 7