**United States Patent**

**Gorin et al.**

**Automated Dialog System and Method**

Inventors: Allen Louis Gorin, Berkeley Heights, NJ (US); Irene Langkilde Geary, Los Angeles, CA (US); Marilyn Ann Walker, Morristown, NJ (US); Jeremy H. Wright, Warren, NJ (US)

Assignee: AT&T Corp., New York, NY (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 90 days.

This patent is subject to a terminal disclaimer.

Appl. No.: 11/219,629

Filed: Sep. 6, 2005

**Related U.S. Application Data**

Continuation of application No. 09/765,444, filed on Jan. 22, 2001, now Pat. No. 7,003,459, and a continuation-in-part of application No. 09/712,194, filed on Nov. 15, 2000, now Pat. No. 6,941,266.

**Int. Cl.**

G10L 15/14 (2006.01)  
G10L 15/22 (2006.01)

**U.S. Cl.**

704/240; 704/257; 704/275

**Field of Classification Search**

704/231, 704/244, 252, 270, 275, 236, 240, 257; 379/88.17, 379/88.18

See application file for complete search history.

**References Cited**

U.S. PATENT DOCUMENTS

4,777,600 A 10/1988 Saito et al.

**Abstract**

This invention concerns a method and system for monitoring an automated dialog system for the automatic recognition of language understanding errors based on a user's input communications in a dialog with the user. The probability of conducting a successful dialog with the user is determined based, at least in part, on understanding data from at least one prior dialog exchange of the dialog.

6 Claims, 6 Drawing Sheets
<table>
<thead>
<tr>
<th>U.S. PATENT DOCUMENTS</th>
<th>OTHER PUBLICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>6,421,655 B1</td>
<td>* cited by examiner</td>
</tr>
<tr>
<td>6,490,698 B1</td>
<td></td>
</tr>
<tr>
<td>6,598,018 B1 *</td>
<td></td>
</tr>
<tr>
<td>6,751,591 B1 *</td>
<td></td>
</tr>
<tr>
<td>6,941,266 B1 *</td>
<td></td>
</tr>
<tr>
<td>7,003,459 B1 *</td>
<td></td>
</tr>
<tr>
<td>7,058,573 B1 *</td>
<td></td>
</tr>
<tr>
<td>7,085,720 B1</td>
<td></td>
</tr>
<tr>
<td>7,127,395 B1 *</td>
<td></td>
</tr>
<tr>
<td>7,158,935 B1 *</td>
<td></td>
</tr>
</tbody>
</table>
START

3000

RECEIVE RECOGNITION AND UNDERSTANDING DATA BASED ON AN INPUT COMMUNICATION FROM A USER

3100

IS THE PROBABILITY OF CORRECTLY UNDERSTANDING THE USER'S INPUT COMMUNICATION ABOVE A FIRST PREDETERMINED THRESHOLD?

3200

YES

STORE DIALOG EXCHANGE INTO DIALOG HISTORY DATABASE

3400

NO

ROUTE FOR HUMAN ASSISTANCE

3300

IS THE PROBABILITY OF CORRECTLY UNDERSTANDING THE USER'S INPUT COMMUNICATION ABOVE A SECOND PREDETERMINED THRESHOLD?

3500

YES

CONDUCT DIALOG USING THE CURRENT DIALOG STRATEGY TO OBTAIN CLARIFICATION FROM THE USER

3600

NO

ADAPT THE DIALOG STRATEGY AND CONDUCT DIALOG USING THE ADAPTED STRATEGY TO OBTAIN CLARIFICATION FROM THE USER

3700

CAN THE USER'S INPUT BE PROCESSED?

3800

YES

3900

END

FIG. 3
START 5000

RECEIVE AN INPUT COMMUNICATION FROM A USER 5100

RECOGNIZE PORTIONS OF THE USER’S INPUT COMMUNICATION 5200

CAN THE USER’S INPUT COMMUNICATION BE CORRECTLY UNDERSTOOD SO THAT THE TASK CAN BE CLASSIFIED? 5300

YES → ROUTE ACCORDING TO TASK OBJECTIVE CLASSIFICATION 5400

NO → IS THE PROBABILITY OF CORRECTLY UNDERSTANDING THE USER’S COMMUNICATION ABOVE A PREDETERMINED THRESHOLD? 5500

YES → STORE DIALOG EXCHANGE INTO DIALOG HISTORY DATABASE 5600

NO → CONDUCT DIALOG TO OBTAIN CLARIFICATION 5700

END 5900

FIG. 5
AUTOMATED DIALOG SYSTEM AND METHOD

CROSS REFERENCE TO RELATED APPLICATIONS


TECHNICAL FIELD

The invention relates to automated systems for communication recognition and understanding.

BACKGROUND OF THE INVENTION

Today there are many automated dialog systems in operation that serve many purposes, such as for customer care. Because they limit human involvement, such systems save millions of dollars in labor costs. Examples of such systems are shown in U.S. Pat. Nos. 5,675,707, 5,800,063, 6,044,337, and 6,173,261, and U.S. patent application Ser. Nos. 08/943,944, filed Oct. 3, 1997, 09/699,494, 09/699,495, and 09/699,496 all filed Oct. 31, 2000, and 09/712,192 and 09/712,194, both filed Nov. 15, 2000, each of which is incorporated by reference herein in its entirety.

While it has recently become possible to build spoken dialog systems that interact with users in real-time in a range of domains, systems that support conversational natural language are still subject to a large number of language understanding errors. Endowing such systems with the ability to reliably distinguish language understanding errors from correctly understood communications might allow them to correct some errors automatically or to interact with users to repair them, thereby improving the system’s overall performance.

SUMMARY OF THE INVENTION

This invention concerns a method and system for monitoring an automated dialog system for the automatic recognition of language understanding errors based on a user’s input communications. The method may include determining whether a probability of understanding the user’s input communication exceeds a first threshold. The method may further operate such that if the first threshold is exceeded, further dialog is conducted with the user. Otherwise, the user may be directed to a human for assistance.

In another possible embodiment, the method operates as above except that if the probability exceeds a second threshold, the second threshold being higher than the first, then further dialog is conducted with the user using the current dialog strategy. However, if the probability falls between a first threshold and a second threshold, the dialog strategy may be adapted in order to improve the chances of conducting a successful dialog with the user.

This process may be cumulative. In particular, the first dialog exchange may be stored in a database. Then, a second dialog exchange is conducted with the user. As a result, a second determination is made as to whether the user’s input communication can be understood and can be conducted based on the stored first exchange and the current second exchanges. This cumulative process may continue using a third and fourth exchange, if necessary.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is described in detail with reference to the following drawings wherein like numerals reference like elements, and wherein:

FIG. 1 is a block diagram of an exemplary NLU monitoring system;
FIG. 2 is a flowchart illustrating an exemplary NLU monitoring process;
FIG. 3 is a flowchart illustrating another exemplary NLU monitoring process using dialog strategy adaptation techniques;
FIG. 4 is a block diagram of an exemplary task classification system using a NLU monitoring system;
FIG. 5 is a flowchart illustrating an exemplary task classification process using a NLU monitoring process; and
FIG. 6 is a flowchart illustrating an exemplary task classification process using a NLU monitoring process having dialog strategy adaptation techniques.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Automated dialog systems promise efficient and natural access to a large variety of information sources and services from any communication device, such as telephones, cellular or mobile phones, computers, videophones, etc. Several hundred systems that support system-initiative dialogs are currently being field-tested and deployed. However, systems that support mixed-initiative, conversational, natural language interaction are still subject to a large number of language understanding errors, which have a large impact on the system’s performance.

This invention concerns monitoring an automated dialog system for the automatic recognition of language understanding errors. The ability to recognize and predict language understanding errors will endow an automated system with the ability to correct some errors automatically or to interact with users to repair them.

As an example of the benefit of error detection, contrast the exemplary dialogs below between human users and first, a conventional dialog system, and then a dialog system including the features of the present invention (the phone numbers, card numbers, and pin numbers in the sample dialogs are artificial):
In contrast, the system of the invention completed the above dialog successfully. Note that after detecting the NLU error in utterance U3, the system of the invention re-prompts the user to repeat the calling card number in utterance S4.

While the above example illustrates the invention in a telephone customer care system, this invention may be applied to any single mode, or multimodal, dialog system. Furthermore, the invention may apply to any automated recognition and understanding system that receives communications from external sources, such as users, customers, service providers, associates, etc. Consequently, the method may operate in conjunction with one or more communication networks, including a telephone network, the Internet, an intranet, Cable TV network, a local area network (LAN), a wireless communication network, etc.

In the discussion below, the term “natural language understanding” is used to describe the understanding of any form of communications received over the networks which may be expressed verbally, nonverbally, multimodally, etc. Examples of nonverbal communications include the use of gestures, body movements, head movements, non-responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes, pointers, stylus, cable set-top box entries, graphical user interface entries, touchscreen entries, etc. Multimodal communications include communications on a plurality of channels, such as aural, visual, etc. However, for ease of discussion, examples and discussions of the method and system of the invention will be discussed below in relation to telephone systems.

FIG. 1 illustrates an exemplary natural language understanding monitoring system 100. The natural language understanding monitoring system 100 includes a natural language understanding (NLU) monitor 180, a dialog manager 190, a training database 165 and a dialog history database 170. The NLU monitor 180 receives recognition data from a recognizer 120 and understanding data from a natural language understanding (NLU) unit 130 that are based on input communications from the user.

The recognizer 120 and the NLU unit 130 are shown as separate units for clarification purposes. However, the functions of the recognizer 120 and the NLU unit 130 may be performed by a single unit within the spirit and scope of this invention.

The recognizer 120 and the NLU unit 130 may operate using one or more of a variety of recognition and understanding algorithms. For example, the recognizer 120 and the NLU unit 130 may use confidence functions to determine whether the user’s input communications have been recognized and understood. The recognition and understanding data from the user’s input communications are also input into the NLU monitor 180. Based on this data, the NLU monitor 180 calculates a probability that the language is understood clearly and this may be used in conjunction with other mechanisms like recognition confidence scores to decide whether and/or how to further process the user’s communication.

As a result, if the user’s input communication can be satisfactorily recognized and understood, the NLU unit 130 routes and/or processes the user’s input communication, which may include the request, comment, etc. However, if the NLU monitor 180 recognizes errors in the understanding of the user’s input communication such that it cannot be satisfactorily recognized and understood, dialog with the user may need to be conducted. This process will be described in greater detail below.

In the natural language understanding monitoring system 100, the dialog history database 170 serves as a database for storing each dialog exchange for a particular dialog. The training database 165 stores NLU errors collected from interactions with human users and models built based on those errors, the NLU features identified from the collected dialogs, and the NLU rules generated from the dialogs and the NLU features. The NLU monitor 180 exploits the training database 165 by using the dialog history stored in the dialog history database 170 to predict whether a NLU error is to occur in the current dialog. While the training database 165 and the dialog history database 170 are shown as separate databases in the exemplary embodiments, the dialog history and training data may be stored in the same database or memory, for example. This database or memory may be stored external or internal to the system.

As discussed below, the NLU monitor 180 of the natural language understanding monitoring system 100 can be trained to improve its ability to detect errors by exploiting language understanding errors collected in interactions with human users and stored in the training database 165. The initial segments of these databases can be used to predict that a problem is likely to occur. The ability to predict language understanding errors will allow the system’s dialog manager 190 to apply more sophisticated strategies to repairing problems, and even perhaps, to prevent them.

Note that the recognizer 120 may be trained to recognize any number of communication symbols, both acoustic and non-acoustic, including grammar fragments, meaningful words, meaningful phrases, meaningful phrase clusters, superwords, morphemes, multimodal signals, etc., using any of the methods known to one skilled in the art including those found in U.S. Pat. Nos. 5,675,707, 5,860,063 and 6,044,337, and U.S. patent application Ser. Nos. 08/943,944, 09/712,192 and 09/712,194, as discussed above.

In describing the invention, three classes of NLU outcomes are distinguished: RCORRECT, a correctly understood utterance; RPARTIAL-MATCH, a partially understood utterance; and RMISMATCH a misunderstood utterance. Experiments were conducted on learning to automatically distinguish these three classes of NLU outcomes in 11,787 spoken utterances collected in a field trials system interacting with live customer traffic.

During the trial, the behaviors of all the system’s 100 components were automatically recorded in a log file and the dialogs were later transcribed by humans and labeled with one or more of the categories representing the intended result/task that the caller was asking the system 100 to perform, on a per communication basis. This label will be referred to as the HUMAN LABEL.

The NLU unit 130 also logged what it believed to be the correct task category. This label will be referred to as the NLU LABEL. One of the focuses of this invention is on the problem of improving the system’s 100 ability to automatically detect when the NLU LABEL is wrong. As mentioned above, this ability would allow the system 100 to make better deci-
The recognizer 120 takes the user’s input communication and produces a transcription. The recognizer 120 features extracted from the corpus were the output of the recognizer 120 (recog), the number of words in the recognizer 120 output (recog-numwords), the duration in seconds of the input to the recognizer 120 (asr-duration), a flag for touchtone input (dtmf-flag), the input modality expected by the recognizer 120 (rg-modality) (one of: none, speech, touchtone, speech+touchtone, touchtone-card, speech+touchtone-card, touchtone-date, speech+touchtone-date, or none-final-prompt), and the grammar used by the recognizer 120 (rg-grammar). A feature called tempo was also calculated by dividing the value of the asr-duration feature by the recog-numwords feature.

The motivation for the recognizer 120 features is that any one of them may have impacted recognition performance with a concomitant effect on language understanding. For example, asr-duration has consistently been found to correlate with incorrect recognition. The name of the grammar (rg-grammar) could also be a predictor of NLU errors since it is well known that the larger the grammar, the more likely a recognizer 120 error is to occur.

One motivation for the temp feature is that it has been found that users tend to slow down their communication (speech, movement, etc.) when the system 100 has misunderstood them. This strategy actually leads to more errors since the recognizer 120 is not trained on this type of communication. The tempo feature may also indicate hesitations, pauses or interruptions, which could also lead to recognizer 120 errors. On the other hand, additional multimodal input such as touchtone (DTMF) in combination with the user’s communication, as encoded by the feature dtmf-flag, might increase the likelihood of understanding, since the touchtone input is unambiguous and will therefore function to constrain language understanding.

The goal of the NLU unit 130 module is to identify which of the possible results, tasks, etc., the user is attempting, and extract from the user’s communication any items of information that are relevant to completing the user’s intended result (e.g., phone numbers are needed for the task dial for me). In this example, fifteen of the features from the NLU unit 130 represent the distribution for each of the 15 possible tasks of the NLU unit 130’s confidence in its belief that the user is attempting that task. Also included is a feature to represent which task has the highest confidence score (top-task), and which task has the second highest confidence score (nexttop-task), as well as the value of the highest confidence score (top-confidence), and the difference in values between the top and next-top-confidence scores (diff-confidence).

Other features represent other aspects of the NLU unit 130 processing of the utterance. The inconsistency feature is an intra-utterance measure of semantic diversity, according to a task model of the domain. Some task classes occur together quite naturally within a single statement or request. For example, the dial for me task is compatible with the collect call task, but is not compatible with the billing credit task.

The salience-coverage feature measures the proportion of the utterance that is covered by the salient grammar fragments. For example, this may include the whole of a phone or card number if it occurs within a fragment. The context-shift feature is an inter-utterance measure of the extent of a shift of context away from the current task focus, caused by the appearance of salient phrases that are incompatible with it, according to a task model of the domain.

In addition, similar to the way we calculated the tempo feature, the salience-coverage and top-confidence features are normalized by dividing them by asr-duration to produce the salience and context features. The motivation for
these NLU features is to make use of information that the NLU unit 130 has as a result of processing the output of recognizer 120 and the current discourse context. For example, for utterances that follow the first utterance, the NLU unit 130 knows what task it believes the user is trying to complete. The context-shift feature incorporates this knowledge of the discourse history, with the motivation that if it appears that the caller has changed his/her mind, then the NLU unit 130 may have misunderstood an utterance. The function of the dialog manager 190 is to take as input the output of the NLU unit 130, decide what task the user is trying to accomplish, decide what the system will say next, and update the discourse history. The dialog manager 190 decides whether it believes there is a single unambiguous task that the user is trying to accomplish, and how to resolve any ambiguity.

Features based on information that the dialog manager 190 logged about its decisions or features representing the ongoing history of the dialog might also be useful predictors of NLU errors. Some of the potentially interesting dialog manager 190 events arise due to low NLU confidence levels computed by the NLU monitor 180 that lead the dialog manager 190 to re-prompt the user or confirm its understanding. A re-prompt might be a variant of the same question that was asked before, or it could include asking the user to choose between two tasks that have been assigned similar confidences by the NLU monitor 180. For example, in the dialog below the system utterance in S3 counts as a re-prompt because it is a variant of the question in utterance S2.

S1: AT&T How may I help you?
U1: I need credit please.
S2: Okay. What is the reason for the credit?
U2: Miss, uh, Miss different area code than I needed.
S3: Sorry, I need to know whether you need credit for a wrong number, bad connection or a call that was cut off. (re-prompt)
U3: It was a wrong number.
S4: What was the number that you dialed?

The features that are extracted from the dialog manager 190 are the task-type label, sys-label, whose set of values include a value to indicate when the system had insufficient information to decide on a specific task-type, the utterance id within the dialog (utt-id), the name of the prompt played to the user (prompt), and whether the type of prompt was a re-prompt (re-prompt), a confirmation (confirm), or a subdialog prompt (a superset of the re-prompt and confirmation prompts (subdialog)). The sys-label feature is intended to capture the fact that some tasks may be harder than others. The utt-id feature is motivated by the idea that the length of the dialog may be important, possibly in combination with other features like sys-label. The different prompt features again are motivated by results indicating that re-prompt are frustrating for users.

The discourse history features included running tallies for the number of re-prompts (num-reprompts), number of confirmation prompts (num-confirms), and number of subdialog prompts (num-subdialogs), that had been played before the utterance currently being processed, as well as running percentages (percent-reprompts, percent-confirms, percent-subdialogs). The use of running tallies and percentages is based on showing that normalized features are more likely to produce generalized predictors.

FIG. 2 is a flowchart of an exemplary natural language understanding monitoring process. The process begins its step 2000 and goes to step 2100 where the NLU monitor 180 receives recognition and understanding data from the recognizer 120 and the NLU unit 130, respectively, based on an input communication from a user. In step 2200, the NLU monitor 180 determines the probability of whether the user's input communication may be correctly understood based on the initial dialog exchange data based on the recognition and understanding data received and the decision rules stored in the training database 165.

If the probability of correctly understanding the user's input communication determined by the NLU monitor 180 does not exceed a predetermined threshold, for example, the NLU monitor 180 signals the dialog manager 190 to route the user to a human for assistance. This situation may be represented by the following relationship:

\[ r_1 + P(\text{understanding}) \]

If the probability of correctly understanding the user's input communication determined by the NLU monitor 180 exceeds the predetermined threshold, for example, the NLU monitor 180 believes that continued dialog may be conducted with the user which may result in successfully routing and/or processing the user's input communication. This situation may be represented by the following relationship:

\[ r_1 - P(\text{understanding}) \]

As result, the process moves to step 2400 where the NLU monitor 180 stores the dialog exchange into dialog history database 170.

In step 2500, the dialog manager 190 conducts further dialog with the user to obtain clarification of the users initial input communication (exchange 1). In step 2600, if the recognizer 120 and NLU unit 130 can recognize and understand the user's second input communication (exchange 2) so that it can be processed, the process goes to step 2700 and ends. However, if after exchange 2 the user's input communication cannot be understood and processed, the process returns to step 2100 where the recognition and understanding data from the user's second input communication is input to the NLU monitor 180 by the recognizer 120 and the NLU 130. Then in step 2200, the NLU monitor 180 determines the probability of correctly understanding the user's input communication based on both the current exchange 2 and previous exchange 1 retrieved from the dialog history database 165. The process that continues as stated above.

FIG. 3 is a flowchart of an exemplary natural language understanding monitoring process that allows the dialog strategy to be adapted between dialog exchanges. Steps 3000, 3100, 3200, 3300, and 3400, operate similarly to steps 2000, 2100, 2200, 2300, and 2400 of FIG. 2, respectively, and will not be further discussed.

However, after step 3400 in which the first dialog exchange is stored in the dialog history database 170, in step 3500, the NLU monitor 180 determines whether the probability of correctly understanding the user's input communication exceeds a second threshold. If the NLU monitor 180 determines that probability of correctly understanding the user's input communication exceeds the second threshold, in step 3600, the dialog manager 190 conducts dialog with the user using the current dialog strategy. This situation may be represented by the following relationship:

\[ r_1 + P(\text{understanding}) \geq r_2 \]

However, if the NLU monitor 180 determines that the probability of correctly understanding the user's input communication does not exceed the second threshold, in step 3700, the dialog manager 190 conducts dialog with the user...
using an adapted (or different) dialog strategy with the user. This situation may be represented by the following relationship:

\[ \text{t1} \neq \text{t2} \]

The process then proceeds to step \textbf{3800}, where determination is made whether the user’s input communication can be processed based on either the current dialog strategy or the adapted dialog strategy opted for in step \textbf{3500}. The process that continues similar to FIG. 2 above, and as such, will not be discussed further.

FIG. 4 illustrates one possible application of the natural language understanding monitoring system \textbf{100}. In particular, FIG. 4 shows an exemplary automated task classification system \textbf{400} that conducts automated dialog with a user. The automated task classification system \textbf{400} includes a classification subsystem \textbf{410}, a labeled training communications database \textbf{450} and a natural language understanding monitoring system \textbf{100}. The classification subsystem \textbf{410} includes a recognizer \textbf{420}, an NLU unit \textbf{430} and a task classification processor \textbf{440}. The natural language understanding monitoring system \textbf{100} includes an NLU monitor \textbf{180}, a dialog manager \textbf{190}, a training database \textbf{165} and a dialog history database \textbf{170}, as discussed above.

The automated task classification system \textbf{400} is based on the notion of task classification and routing. In the automated task classification system \textbf{400}, services that the user can access are classified into a plurality of categories by a task classification processor or alternatively, to a category called other for calls that cannot be automated and must be transferred to a human operator. Each category describes a different task, such as person-to-person dialing, or receiving credit for a misdialed number.

The task classification processor \textbf{440} determines which task the user is requesting on the basis of the NLU unit’s \textbf{430} understanding of recognized language segments, stored in the labeled training communications database \textbf{450}, and recognized by the recognizer \textbf{420}. The NLU \textbf{430} attempts to interpret the user’s response to, for example, the open-ended system greeting “Hi, how may I help you?” On the other hand, once the response has been recognized and understood, task classification processor \textbf{440} determines the task, and the information needed for completing the caller’s request is obtained using the dialog manager \textbf{190}. The dialog manager \textbf{190} uses sub-modules that are specific for each task.

While the task classification processor \textbf{440} and the NLU unit are shown as separate units in FIG. 4, their functions may be performed by a single unit within the spirit and scope of the invention.

FIG. 5 is a flowchart of a possible automated task classification process using the natural language understanding monitoring system \textbf{100} of the invention. The process begins at step \textbf{5000} and proceeds to step \textbf{5100} where an input communication is received by the recognizer \textbf{420}. At step \textbf{5200}, the recognizer \textbf{420} attempts to recognize portions of the user’s input communication, including grammatical words/phrases/symbols, morphemes, actions, gestures, or any other communication signal.

At step \textbf{5300}, based on the understanding received from the NLU unit \textbf{430}, the NLU monitor \textbf{180} determines whether the user’s input communication can be correctly understood so that the task can be classified. If the task can be classified, then in step \textbf{5400}, the task classification processor \textbf{440} routes the user’s request according to the task classification, and the process proceeds to step \textbf{5500} and ends.

However, in step \textbf{5300}, if the NLU monitor \textbf{180} determines that the task classification processor \textbf{440} cannot classify the user’s request, in step \textbf{5500}, the NLU monitor \textbf{180} determines whether the probability of correctly understanding the user’s input communication exists above a predetermined threshold. In this iteration, the NLU monitor \textbf{180} is using only the first exchange (exchange 1). The NLU monitor \textbf{180} uses the classification model stored in the dialog training database \textbf{165} to determine whether the probability of correctly understanding the user’s input communication exceeds the predetermined threshold. If the NLU monitor \textbf{180} determines that the probability of correctly understanding the user’s input communication does not exceed the threshold, then in step \textbf{5800} the user is routed to a human for assistance. The process then goes to step \textbf{5900} and ends.

If, in step \textbf{5500}, the NLU monitor \textbf{180} determines that the probability of correctly understanding the user’s input communication is above the predetermined threshold based on the stored classification model, in step \textbf{5600} the first dialog exchange is stored in the dialog history database \textbf{170}. Then, in step \textbf{5700}, the dialog manager \textbf{190} conducts dialog with the user to obtain further clarification of user’s request. The process returns to step \textbf{5200} wherein the recognizer \textbf{420} attempts to recognize portions of the second exchange with the user.

Steps \textbf{5300} and \textbf{5400} are performed as before using the second exchange.

However, in step \textbf{5400}, if the task classification processor \textbf{440} determines that the task cannot be classified, the dialog predictor \textbf{180} gauges the probability of conducting successful dialog above the predetermined threshold based on both exchanges 1 and 2 and exchange 1 stored in the dialog history database \textbf{170}. If the probability of conducting a successful dialog does not exceed the predetermined threshold based on exchanges 1 and 2, in step \textbf{5800}, the task classification processor \textbf{440} is instructed by the dialog manager \textbf{190} to route the user to a human assistant.

On the other hand, in step \textbf{5500}, if the NLU monitor \textbf{180} determines that the probability of correctly understanding the user’s input communication exceeds the predetermined threshold, the second dialog exchange may be stored in step \textbf{5600} and further dialog conducted with the user in step \textbf{5700}. The process would continue using the current and previous exchanges with the user until the task is completed or the user is routed to a human assistant.

FIG. 6 is a flowchart of a possible automated task classification process using the natural language understanding monitoring system \textbf{100} with the ability to adapt dialog strategy in an effort to conduct successful dialog with the user. Steps \textbf{6000}, \textbf{6050}, \textbf{6100}, \textbf{6150}, \textbf{6250}, \textbf{6300}, \textbf{6350}, and \textbf{6400} are performed similarly to steps \textbf{5000}, \textbf{5100}, \textbf{5200}, \textbf{5300}, \textbf{5400}, \textbf{5500}, \textbf{5600}, and \textbf{5700} of FIG. 5, respectively, and will not be discussed further.

However, after step \textbf{6400} in FIG. 6 in which the first dialog exchange is stored in the dialog history database \textbf{170}, in step \textbf{6450}, the NLU monitor \textbf{180} determines whether the probability of correctly understanding the user’s input communication exceeds a second threshold. If the NLU monitor \textbf{180} determines that the probability of correctly understanding the user’s input communication exceeds the second threshold, in step \textbf{6500}, the dialog manager \textbf{190} conducts dialog with the user using the current dialog strategy.

However, if the NLU monitor \textbf{180} determines that the probability of correctly understanding the user’s input communication does not exceed the second threshold, in step \textbf{6500}, the dialog manager \textbf{190} conducts dialog with the user using an adapted (or different) dialog strategy with the user.

The process then proceeds back to step \textbf{6500}, where the recognizer \textbf{120} receives the user’s input communication based on either the current dialog strategy or the adapted
dialog strategy opted for in step 6450. The process that continues similar to FIG. 5 above, and as such, will not be discussed further.

The output of each experiment is a classification model learned from the training data that is stored in the training database 165. The model is evaluated in several ways. First, multiple models are trained using different feature sets extracted from different system 100 components in order to determine which feature sets are having the largest impact on performance. Second, for each feature set, the error rates of the learned classification models are estimated using ten-fold cross-validation, by training on a random 10,608 utterances and testing on a random 1,179 utterances 10 successive times. Third, precision, recall and the confusion matrix are recorded in the classifier, and then trained on all the features tested on a random held-out 20% test set. Fourth, for the classifier trained on all the features, the extent to which the error can be minimized on the error classes RMISMATCH and RPARTIAL-MATCH is examined by manipulating the rule-learning program’s loss ratio parameter. Finally, the results of training other learners are compared on the same dataset with several of the feature sets. The overall accuracy results for detecting NLU errors using the rule-learning program are summarized below (SE=Standard Error).

<table>
<thead>
<tr>
<th>Features Used</th>
<th>Accuracy (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASELINE (majority class)</td>
<td>63.47% (0.38)</td>
</tr>
<tr>
<td>ALL</td>
<td>86.16% (0.38)</td>
</tr>
<tr>
<td>NLU UNIT ONLY</td>
<td>84.80% (0.38)</td>
</tr>
<tr>
<td>RECOGNIZER+DISCOURSE</td>
<td>80.97% (0.26)</td>
</tr>
<tr>
<td>RECOGNIZER ONLY</td>
<td>78.89% (0.27)</td>
</tr>
<tr>
<td>DISCOURSE ONLY</td>
<td>71.97% (0.40)</td>
</tr>
</tbody>
</table>

The first line of the above table represents the accuracy from always guessing the majority class (RCCORRECT); this is the BASELINE against which all the other results should be compared. The first row labeled ALL, shows the accuracy based on using all the features available from the system 100 components. This classifier can identify NLU errors 23% better than the baseline. The second row of the table, NLU ONLY, shows that the classifier based only on the NLU unit 130 features performs statistically as well as the classifier based on all the features. The third row of the table, RECOGNIZER+DISCOURSE shows that combining the RECOGNIZER features with the DISCOURSE features produces a significant increase in accuracy over the use of recognizer 120 features alone, which however still performs worse than the NLU unit 130 features on their own. The last two rows, RECOGNIZER ONLY and DISCOURSE ONLY, indicate that it is possible to do significantly better than the baseline using only the features from the recognizer 120 or from the dialog manager 190 and the discourse history, but these features on their own cannot do as well at predicting NLU accuracy as the NLU unit’s 130 own features based on its own calculations.

Using ten-fold cross-validation, the resulting NLU error classifier can correctly identify whether an utterance is an NLU error 83% of the time, an improvement of 23% over the majority class baseline. In addition, the most important features are those that the NLU unit 130 can compute, suggesting that it will be straightforward to integrate the NLU monitor 180 into the NLU unit 130 of the system 100.

Further results are discussed in Walker et al., “Using Natural Language Processing and Discourse Features to Identify Understanding Errors in a Spoken Dialogue System”, Inter-

As shown in FIGS. 1 and 4, the method of this invention may be implemented using one or more programmed processors. However, method can also be implemented on a general-purpose or a special purpose computer, a programmed microprocessor or microcontroller, peripheral integrated circuit elements, an application-specific integrated circuit (ASIC) or other integrated circuits, hardware/electronic logic circuits, such as a discrete element circuit, a programmable logic device, such as a PLD, PLA, FPGA, or PAL, or the like. In general, any device on which the finite state machine capable of implementing the flowcharts shown in FIGS. 2, 3, 5 and 6 can be used to implement the task classification system and natural language understanding monitoring functions of this invention.

While the invention has been described with reference to the above embodiments, it is to be understood that these embodiments are purely exemplary in nature. Thus, the invention is not restricted to the particular forms shown in the foregoing embodiments. Various modifications and alterations can be made thereto without departing from the spirit and scope of the invention.

The invention claimed is:

1. A method for use in an automated dialog system in which a dialog comprising dialog exchanges is conducted between the system and a user, at least ones of the dialog exchanges of the dialog including an input communication from the user, the input communication including one or more communication symbols, the method comprising recognizing at least one of the communication symbols of the input communication of a present dialog exchange to determine the input communication, and determining a probability of conducting a successful dialog with the user, a successful dialog being a dialog exchange between the automated dialog system and the user that results in processing of input communications of the user to a stage consistent with the user’s intent, said probability being determined based, at least in part, on understanding data from at least one prior dialog exchange of the dialog, said understanding data being data indicative of a probability that the input communication of said at least one prior dialog exchange was correctly understood, wherein said one or more communication symbols includes at least one of words, phrases, morphemes, and gestures,

wherein said understanding data for a particular dialog exchange of the dialog includes at least one confidence score, said confidence score being a measure of a confidence that a particular task associated with said confidence score was a task that the user was requesting the system to perform in said particular dialog exchange, wherein said at least one confidence score is a top-confidence score, said top-confidence score being the highest confidence score among a plurality of confidence scores, each of the plurality of confidence scores being a measure of a confidence that a particular respective task associated with said each confidence score was a task that the user was requesting the system to perform in said particular dialog exchange,

wherein said understanding data for a particular dialog exchange of the dialog includes a confidencetime feature, said confidencetime feature being said top-confidence score normalized by the duration of the input communication.

2. A method for use in an automated dialog system in which a dialog comprising dialog exchanges is conducted between
the system and a user, at least ones of the dialog exchanges of
the dialog including an input communication from the user,
the input communication including one or more communica-
tion symbols, the method comprising
recognizing at least one of the communication symbols of
the input communication of a present dialog exchange to
determine the input communication, and
determining a probability of conducting a successful dia-
log with the user, a successful dialog being a dialog
exchange between the automated dialog system and the
user that results in processing of input communications
of the user to a stage consistent with the user’s intent,
said probability being determined based, at least in part,
on understanding data from at least one prior dialog
exchange of the dialog, said understanding data being
data indicative of a probability that the input communi-
cation of said at least one prior dialog exchange was
correctly understood,
wherein said one or more communication symbols
includes at least one of words, phrases, morphemes, and
gestures,
wherein the input communication is an utterance and
wherein said understanding data for a particular dialog
exchange of the dialog includes a salience-coverage fea-
ture, said salience-coverage feature being a measure of
the proportion of the utterance that is covered by salient
grammar fragments in said utterance,
and wherein said understanding data for a particular dialog
exchange of the dialog includes a salpertime feature,
said salpertime feature being said salience-coverage fea-
ture normalized by the duration of the input communi-
cation.
3. A method for use by an automated dialog system, the
method comprising
determining a probability that the system will correctly
understand a user’s intent based on a dialog between the
user and the system, said probability being determined
based, at least in part, on understanding data of a) an
input communication of a present dialog exchange
between said system and said user and b) an input commu-
nication of one or more dialog exchanges between
said system and said user that were prior to said present
dialog exchange,
each said understanding data being data associated with a
respectively input communication from the user and being
indicative of the probability that the system has correctly
understood the associated input communication
wherein said user’s intent is to ask the system to perform a
particular task,
wherein said understanding data for a particular input commu-
nication includes at least one confidence score, said
certainty score being a measure of a certainty that a particu-
lar task associated with said confidence score was
a task that the user was requesting the system to perform
in said particular input communication,
wherein said at least one certainty score is a top-certainty score,
said top-certainty score being the highest certainty score
among a plurality of certainty scores, each of the plurality of certainty scores being a measure of a certainty that a particular respective task associated with said certainty score was a task that the user was requesting the system to perform in said particular input communication,
and wherein said understanding data for a particular input commu-
nication includes a salpertime feature, said
certainty score being said top-certainty score normalized by the duration of said particular input communication.
4. A method for use by an automated dialog system, the
method comprising
determining a probability that the system will correctly
understand a user’s intent based on a dialog between the
user and the system, said probability being determined
based, at least in part, on understanding data of a) an
input communication of a present dialog exchange
between said system and said user and b) an input commu-
nication of one or more dialog exchanges between
said system and said user that were prior to said present
dialog exchange,
each said understanding data being data associated with a
respectively input communication from the user and being
indicative of the probability that the system has correctly
understood the associated input communication,
wherein said user’s intent is to ask the system to perform a
particular task,
wherein at least one of said input communications is an
utterance and wherein said understanding data for that
input communication includes a salience-coverage fea-
ture, said salience-coverage feature being a measure of
the proportion of the utterance that is covered by salient
grammar fragments in said utterance,
and wherein said understanding data for said at least one of
said input communications includes a salpertime feature,
said salpertime feature being said salience-coverage
feature normalized by the duration of said at least one of the input communications.
5. A method for use in an automated dialog system that
conducts a dialog including input communications from a
user, the automated dialog system being adapted to recognize
the input communications, the method comprising
determining a probability that the dialog being successful,
the dialog being successful when it results in the auto-
nated dialog system correctly understanding the
intended natural language meaning of the input communica-
tions from the user, said probability being deter-
mined based, at least in part, on natural language under-
standing data indicative of whether each of a present one
of the input communications and at least one prior one of
the input communication have been correctly under-
stood,
wherein said natural language understanding data for said
each input communication includes at least one confi-
dence score, said confidence score being a measure of a
confidence that a particular task associated with said confidence score was a task that the user was requesting the system to perform in said each input communication,
wherein said confidence score is a top-confidence score,
said top-confidence score being the highest confidence
score among a plurality of confidence scores associated
with said each input communication, each of the plural-
ity of confidence scores being a measure of a confidence that a particular respective task associated with said confidence score was a task that the user was requesting the system to perform in said each input communication,
and wherein said natural language understanding data for
said each input communication includes a salpertime feature, said salpertime feature being said top-confidence score normalized by the duration of said each input communication.
6. A method for use in an automated dialog system that
conducts a dialog including input communications from a
user, the automated dialog system being adapted to recognize the input communications, the method comprising determining a probability of the dialog being successful, the dialog being successful when it results in the automated dialog system correctly understanding the intended natural language meaning of the input communications from the user, said probability being determined based, at least in part, on natural language understanding data indicative of whether each of a present one of the input communications and at least one prior one of the input communication have been correctly understood, wherein at least one of said each input communication is an utterance and wherein said natural language understand-

ing data for that input communication includes a salience-coverage feature, said salience-coverage feature being a measure of the proportion of the utterance that is covered by salient grammar fragments in said utterance, and wherein said natural language understanding data for said each input communication includes a salpertime feature, said salpertime feature being said salience-coverage feature normalized by the duration of said each input communication.