実時間音声対話システムTOSBURGの開発 (3) 対話処理

6N−7

新地 秀昭* 貞本 洋一** 坪井 宏之*** 竹林 洋一**

*東芝ソフトウェアエンジニアリング(株) **(株)東芝 総合研究所 ***(株)東芝 関西研究所

1.はじめに

計算機との音声による自然なコミュニケーションを実現するためには、音声理解、応答生成およびそれらを統合するための対話処理が重要である[1][2][3][4]。従来の音声認識システムでは、ユーザはシステムの要求にしたがって型通りに入力を行なう必要があった。しかし、話し言葉では、不要語(間投詞、言い淀みなど)、語順の入れ替え、省略が頻繁に見られ、対話の形態も多様であるため、計算機との自然な対話を実現するためには、できるだけ制約の少ない自由発話の理解を行なうとともに、認識処理における曖昧性や誤認識などに対しても柔軟に対処できるような対話制御を行う必要がある。 本稿では、限定したタスクで、利用者の発話に極力制限を設けない不特定ユーザ向の実時間対話システムとして開発したTOSBURG(Task-Oriented dialogue System Based on speech Understanding and Response Generation)のユーザ主導型の対話処理について述べる。

2.対話のモデル

音声による対話システムを構築する際には、認識誤りを含む多様な発話を受入れ、タスクや対話の流れに即した発話理解を行い、不完全な部分については、その内容や対話の前後関係などに基づき、確認や質問などの応答を返すことによりユーザを導く必要がある。 TOSBURGの対話制御部では、ユーザ主導型の対話を実現するため、ユーザの発話を理解する状態(ユーザ状態)とタスクを管理して応答を生成する状態(システム状態)に分け、図1のように対話を両者間の遷移としてモデル化した。 ユーザ状態では、直前のシステムの応答や対話の履歴情報などを参照してユーザの発話を理解し、それに対応したシステムの状態に遷移する。一方、システム状態では、理解した発話内容にしたがって対話の履歴情報を更新し、その確認もしくは不完全な部分についての質問などの応答を生成することにより対話を進行し、その応答に対応したユーザの状態に遷移する。 上記のモデルにより状況に応じたincrementalな対話音声理解と応答の生成が可能となる。 また、複雑な対話制御に関する知識を簡潔に記述できるので変更や拡張などが容易となる。 本システムでは"ハンバーガーショップでの注文"をタスクとしており、図2に示した対話例のように、ユーザの発話は、品目の注文、追加、削除、訂正およびシステムの確認に対する肯定や否定などに分類できる。システムは、それらの確認のための応答を出力し対話を進めている。 なお、対話制御部では、ユーザの発話およびシステムの応答は文字面ではなくその意味を抽象化したフレーム形式で表現する。 以下、これらの意味表現を用いたユーザ状態における発話理解とシステム状態における応答生成について述べる。

3.ユーザ状態における対話音声理解

対話処理部では、音声理解部から渡されたユーザの発話に対する複数の入力意味表現候補について、ユーザ状態に依存

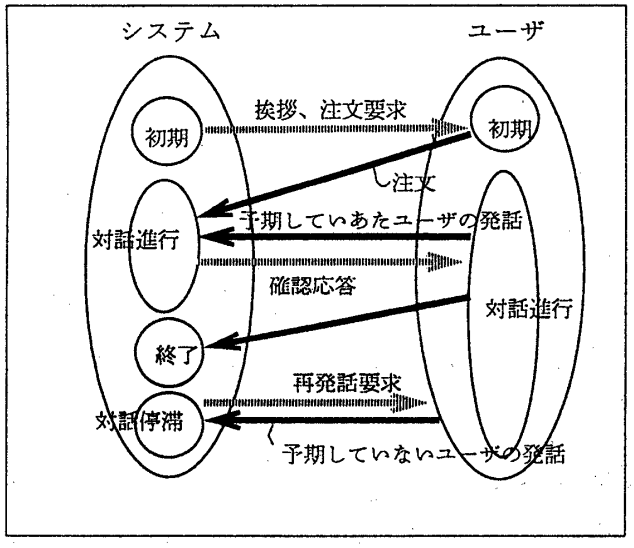

図1. 対話のモデル

1)システム: いらっしゃいませ、ご注文をどうぞ。
2)ユーザ: ハンバーガーと、コーヒーを2つずつ下さい。
3)システム: ご注文はハンバーガーを2つ、コーヒーを2つですね。
4)ユーザ: はい。
5)システム: 以上でよろしいですか。
6)ユーザ: あと、ポテトを下さい。
7)システム: ポテトの小を1つ追加ですね。
8)ユーザ: いいえ、2つです。
9)システム: ポテトの小は2つですか。
10)ユーザ: ええ、そうです。
11)システム: ご注文は、ハンバーガーを2つ、ポテトの小を2つ、コーヒーを2つですね。
12)ユーザ: はい。
13)システム: ありがとうございました。

図2. 対話例

Development of Real-Time Speech Dialogue System TOSBURG (3) Dialogue control
Hideaki SHINCHI*, Yoichi SADAMOTO**, Hiroyuki TSUBOI***, Yoichi TAKEBAYASHI**
* Toshiba Software engineering, ** Toshiba R&D Center, *** Toshiba Kansai Research Lab.

した解析を行い、直前のシステムの応答と対話の履歴情報との意味的な整合性を調べ、その時点で最も妥当な候補を選択する。

3.1.省略部分の補完

対話中の語の省略に対処するため、本システムでは候補の評価を行なう前に省略された部分(品名、サイズ、個数)を直前のシステムの応答を参照して補完する。サイズ、個数については、焦点や対話の履歴情報などを考慮し省略部分を補完するが、適当なものが無ければ、default値を設定する。図3は図2の対話例の8)に対する処理を示しており、音声理解部から2つの入力意味表現候補が入力されている。ここで、候補1では品名とサイズが省略されているが、直前のシステムの応答である出力意味表現を参照することにより、その品名"ポテト"とそのサイズ"中"が補完される。候補2については省略がないためこの処理は行なわない。

3.2.評価・選択

音声理解部で求められた入力意味表現候補はキーワードの尤度や構文的制約によってスコアリングされているが、ユーザ状態ではさらに対話の流れに照し合せた意味的制約を用いてスコアリングを行なう。図3の入力意味表現候補2は、品目"コーラの大"の削除を表しているが、直前のシステム応答の出力意味表現は削除の対象となる品目ついての応答ではなく、さらに、削除されるべき品目はまだ注文されていないのでスコアを減じる。候補1には、意味的に非整合な情報は含まれていないのでスコアの変更は行なわない。こうして、各候補に対して意味的な整合性を調べてスコアリングを行ない、最後に各候補のスコアを比較することにより1つの入力意味表現を選択する。

4.システム状態における応答生成

システム状態では、理解した発話内容にしたがって対話の履歴情報を更新し、システムが発話を正しく理解していることを伝えるための応答を出力する。また、ユーザの発話が理解できなかったり、理解した内容が意味的に受け入れられない場合は、もう一度発話を要求する応答を返すことによって対話を進める。応答の内容は、入力意味表現と同じくフレーム型式で表現し、入力意味表現の確信度をつけて、これを応答生成・出力部に出力する。上述した出力意味表現を介して、応答文の生成や応答音声のイントネーション、画面に表示される店員の表情などが決まる。

5.むすび

task-orientedな不特定話者向の実時間音声対話システムTOSBURGにおける対話処理について述べた。対話の流れをユーザ状態とシステム状態とに分けて記述した対話モデルにより、状況に応じたincrementalな対話音声理解と応答生成を行なうことができる。今後、本システムを用いて評価実験を行なうとともに、より柔軟なモデル化を行なう予定である。

参考文献

[1]B.J.Grosz,C.L.Sidner,"The Structure of Discourse Structure",Computational Linguistics,1986

[2]Allen,J.F.,"Recognizing Intention from Natural language Utterances",Computational Model of Discourse ,p107-166,1983

[3]小林 哲則,白井 克彦:"ネットワークモデルによる会話音声理解における焦点の表現法",音声研究会資料 S85-15,,1985-6

[4]速水 悟,伊藤 克亘,田中 和世:"音声対話システムの構築とそれを用いた会話音声収集",音声研究会資料,SP91-101,1991-12

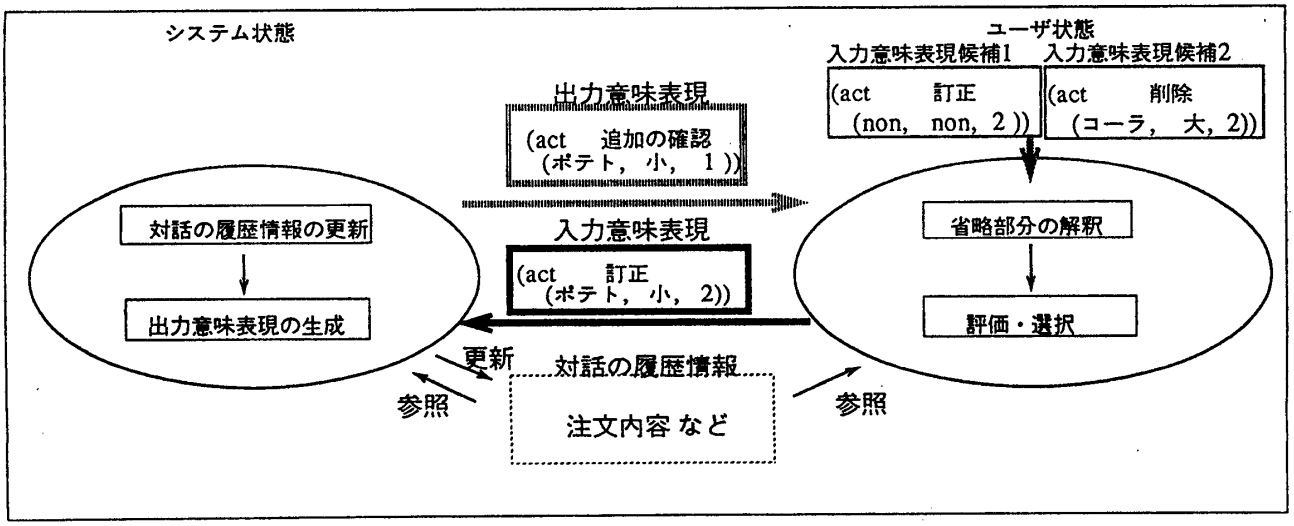

図3 対話制御部での処理