マルチモーダルインタフェースを備えた観光案内対話システム

傳田 明弘　伊藤 敏彦　中川 聖一

豊橋技術科学大学 情報工学系
〒441 愛知県豊橋市天伯町雲雀ヶ丘 1-1

1 はじめに

音声認識技術、及び、これを支援する自然言語処理技術の向上により、ユーザがより自然な言い回しで対話を行える音声対話システムが実現されてきている。また、音声対話システムの対話を支援する目的で、「音声入出力」以外に「ポインティングジェスチャ入力」等の複数の入出力「手段」（モダリティ）を同時に用いるマルチモーダルインタフェースの研究も近年盛んに行われるようになってきている [1, 2]。

我々の研究室では「富士山観光案内日本語音声対話システム」[3, 4] の開発を行なってきた。しかし、マンマシンインタフェースを音声のみで提供していた従来のシステムでは、システムからの応答が音声のみで行われるために、そこに情報の欠落が生じたり対話状況が不透明になってしまうといった問題が生じ、ユーザに不安や負担を与えることになりかねなかった。

そこで、上記の問題点の解消を目的として、現在のシステムでは、「システムとの対話の途中経過表示」や「タッチ入力、及び、指示語（指示詞及び指示代名詞）を含んだユーザ発話の許可」といった、マルチモーダルインタフェースの実現を行なっている [5, 6]。このようなマルチモーダルインタフェースによって、システムとの対話のバリエーションも増え、より自然で内容の豊かな対話が行なえるようになることが期待される。

2 従来の「富士山観光案内日本語音声対話システム」

「富士山観光案内日本語音声対話システム」[3, 4] は、富士山周辺の観光案内をタスクとしており、ユーザの発声する音声を入力とし、その発話内容に対する観光案内を合成音声で応答する。現在のシステムでは、普段我々が使用しているような話し言葉に近い『自然な発話 (Spontaneous Speech)』を理解することが可能になっている。ここでいう『自然な発話 (Spontaneous Speech)』とは、発話文中に「関投詞」「未知語」「助詞落ち」「言い淀み・言い直し」「倒置」といった話し言葉特有の現象を含んだ発話のことである。

従来の「富士山観光案内音声対話システム」は、「入力音声認識部」「対話理解・管理部」及び「応答音声合成部」の 3 つの部分から構成されている。

認識に用いる HMM（Hidden Markov Model）には、日本語の 113 音節を単位とする音節 HMM で、5 状態 4 出力分布の単一連続分布を持つ、離散継続時間制御 HMM（DDCHMM）を用いている[1]。更に、HMM は話者適応化を行なって、認識率の向上をはかっている。この音節 HMM 、文脈自由文法の構文解析法、音声の「取り込み」「分析」「認識」を同時に行なう並列化アルゴリズム、及び、One Pass Viterbi サーチアルゴリズムに基づいたフレーム同期型の連続音声認識の統合アルゴリズムを基礎として、ユーザの発話を認識する「入力音声認識部」は構成されている。更に「間投詞」や「言い淀み・言い直し」の部分には、未知語処理に基づいた処理を施している。文脈自由文法は自然な対話音声を認識するために、助詞落ちや倒置を含む文も受理できるように記述している。

ユーザの発話は「入力音声認識部」で認識され、5-best の認識結果が、「対話理解・管理部」に送られる。この内第 1 位の認識結果のみを文字列に変換し、変換した文字列に対して「形態素解析」「文節解析」「構文解析」「意味解析」「文脈解析」を行い、続いて、富士山周辺の観光地データベースの検索を行なうことによって、応答文の文字列を生成する。「構文解析」及び「意味解析」においては、助詞落ち、助詞誤り、倒置に対応するためにいくつかのヒューリスティックスを用いて解析を行っている。

「応答音声合成部」は、対話システムが生成する応答文の文字列をワークステーション上で動作する音声合成サーバに送り、音声を出力している。

3 マルチモーダルインタフェースの実現

3.1 従来のシステムの問題点

従来のシステムでは、マンマシンインターフェースとして音声のみを用いているため、以下のような問題点が生じてくる。

1. システムからの応答文が多くの内容を含んでいる場合、システムの発話は長くなり、応答内容の一部を聞き逃す可能性がある。
2. 応答された観光地名、施設名等の漢字表記がわからないことがある。
3. ユーザはシステムと対話を行う際に、システムから得られる情報をメモを取る等の手段で記録しながら、対話を進めていかなくてはならない。

このように、対話によって得られる情報の一部が欠落してしまうことや、ユーザにメモを取らせること、対話状況が不透明であることは、ユーザに不安、若しくは、負担を与えることになりかねない。

3.2 ディスプレイ上への情報表示

システムからの応答や過去の対話で得られた情報を記憶し、ユーザが必要とする情報を、対話の途中経過表示として、以下の 4 種類の手段によって画面上に表示する。

[1]デモ用・オンラインによる評価用のシステムでは、システムの応答時間短縮を優先しているため、この離散継続時間制御は使用していない。

- 現在の対話内容に対応する場所の地図 (富士山、河口湖、山中湖、西湖、精進湖 & 本栖湖、のいずれか) 及び現在のトピックを表示。
- 現在の対話内容の対象が観光地や観光施設である場合、その場所の静止画像を表示。
- システムからの応答文が多く (今回の改良においては 5 個以上) の項目を含んでいる場合に、これらをメニューとしても表示。
- システムとの対話から得られた情報の内、各観光施設の (名称、種類、料金や食事、駐車場の有無といったその他の) 情報は、対話履歴として随時表示。

これらをディスプレイ上に表示した画面の例を図 1に示す。

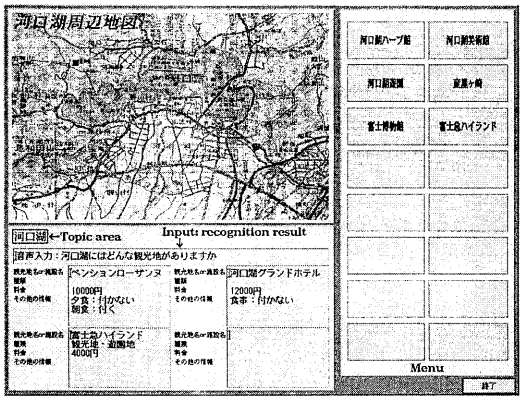

図 1: 画面表示の例
(左上：地図、左中央：現在のトピックと入力音声の認識結果、左下：対話履歴、右：応答文のメニュー)

3.3 タッチスクリーン入力法

ユーザがシステムからの応答文の一部を聞き逃してしまった場合にも、以降の対話で聞き逃した観光地名等を対象とする質問を行なえるようにメニューが表示される。このメニュー内の項目、若しくは、地図上の位置や地名等に指を触れながら (タッチ入力) 、これを指示語で言い表すことによって、システムとの対話を行えるタッチスクリーン入力法を実現している。

4 デモシステムの仕様と対話例

オンラインデモシステムの仕様を以下に、システムとの対話例を図 2 に示す。

- 計算機システムの構成 :
  HP Model 735/125 (音声の取り込み及び分析)
  HP Model J200 (音声の認識)
  HP Model 712/80 (形態素解析以外の言語処理 ; AKCL で記述されている)
  SUN SPARC station 10 (形態素解析 , 音声合成 , 画像表示 , タッチ入力処理)
- 語彙数 : 245
- パープレキシティ : 74
- ビーム幅 : 5 (オフライン時には 20 を使用)
- 継続時間制御なし、ケプストラムの回帰係数未使用 (オフライン時には、両方とも使用)
- 音声認識時間 : リアルタイムの約 3 ～ 6 倍 (オフライン時には、12 ～23 倍)
- 言語処理時間 : 約 0.4 ～ 1.4 秒 / 文
- 音声認識・理解精度 : 文認識率 (但し、助詞誤りは無視している) 約 41% , 文理解率 約 39% (オフライン時にはそれぞれ 約 62%, 60%)
- タスク (1 泊 2 日の旅行計画) 達成率 : 100%

SYS: 富士山観光案内システムです。ご用件をお願いします。
USR: えーと、河口湖にはどんな観光地がありますか。
SYS: 河口湖には河口湖美術館や河口湖遊園や産屋ヶ崎や富士博物館や富士急ハイランドがあります。
(河口湖の地図、及び、5 箇所の観光地名がメニューとして表示される)
USR: あのー、いくらかかるんでしょうか。富士急ハイランドの入場料って。
SYS: 4000 円かかります。
(富士急ハイランドの画像が表示される)
USR: (地図上の湖の位置に指で触れて、)
この周辺にはどんな宿泊施設があるんですか。
SYS: 河口湖にはホテルやペンションや旅館があります。
USR: どんなペンションがありますか。
SYS: 河口湖にはペンションローザンヌやペンションクレヨンがあります。

図 2: 対話例

5 おわりに

今後は、このシステムの評価を行なっていく。また、画像や音声の組合せで実現されるマルチモーダルインタフェースが、人間に与える心理的効果を考察し、音声対話システムにとっての「より使い勝手のよい」マンマシンインタフェースをどのように構築していくかを検討していくつもりである。

参考文献

[1] 竹林洋一 :「音声自由対話システム TOSBURG II － ユーザ中心のマルチモーダルインタフェースの実現に向けて －」, 電子情報通信学会論文誌 , VOL.J77-D-II, No.8, pp.1417-1428 (1994).

[2] 中川聖一, 張建新 :「音声と直指操作による入力インターフェース」, 電気学会論文誌, Vol.114-C, No.10, pp.1009-1017(1994)

[3] M.Yamamoto, S.Kobayashi, Y.Moriya, S.Nakagawa: "A Spoken dialog system with verfication and clarification queries", IEICE Trans., Vol.E76-D, No.1, pp.84-94 (1993)

[4] 山本, 伊藤, 肥田野, 中川 :「人間の理解手法を用いたロバストな音声対話システム」, 情報処理学会論文誌, VOL.37, No.4, pp.471-482(1996).

[5] 傳田 明弘 , 中川 聖一 :「日本語音声による観光案内システムのマルチモーダルインターフェイス化」, 情報処理学会第 52 回全国大会 (2) , 4D-3, pp.167-168 (1996).

[6] 傳田 , 伊藤 , 中川 :「マルチモーダルインタフェースを備えた観光案内対話システムの評価」, 人工知能学会第 10 回全国大会論文集 , 15-09, pp.431-434 (1996).