

# 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

4D-3

傳田 明弘 中川 聖一  
豊橋技術科学大学 情報工学系

## 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ーフェイ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]。

我々の研究室で開発を行ってきた「富士山の観光案内」をタスクとする音声対話システムでは、マンマシンインターフェイスとしては主に音声のみを用いていた。しかし、システムからの返答が音声のみで行われる場合には、返答文が多くの内容を含んでいると、返答内容の一部をユーザが聞き逃してしまう等の問題が生じるために、ユーザに不安や負担を与えることになりかねなかった。

そこで本研究では、ユーザに不安や負担を与えないことを目的とした、システムとの対話の途中経過表示や、タッチ入力&指示代名詞を含んだ発話の実現といった、システムのマルチモーダルインターフェイス化を行った。このようなマルチモーダルインターフェイスの付加により、システムとの対話のバリエーションも増え、より自然な対話が行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」

従来の「富士山観光案内日本語音声対話システム」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 [3, 4]。

ユーザの発話を認識する音声認識部は、HMMを音節のモデルとして用い、文脈自由文法の構文解析法とフレーム同期型連続音声認識の統合アルゴリズムを基礎としたものである。更に間投詞や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。

対話理解・管理部は、ユーザ発話の認識結果の形態素解析、文節解析、構文解析、意味解析、文脈解析を行った後、富士山周辺の観光地データベースの検索を行って、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、これに基づいて解析を行っている。

Development of the multimodal interface for a sightseeing guidance system based on spoken Japanese  
Akihiro Denda and Seiichi Nakagawa  
Toyohashi University of Technology

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 システムのマルチモーダルインターフェイス化

### 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イスとしては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。

- システムからの返答文が多くの内容を含んでいる場合、システムの発話が長くなり、返答内容の一部を聞き逃す可能性がある。
- 返答された観光地名、施設名等をどのような漢字で表記したら良いか、わからないことがある。
- 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なくてはならない。

例えば、「富士山にはどんな観光地がありますか」というユーザの発話に対して、実に16箇所もの観光地名が含まれた返答がなされる。このような返答を音声で、しかも途中で止まること無く返答されても、ユーザはその全てを聞き取りきれない可能性が高い。

このような例の他に、対話によって得られる情報の一部が欠落してしまったり、対話状況が不透明であることは、ユーザに不安、もしくは、負担を与えることになりかねない。そこで、ユーザに余計な不安や負担を与えないようにすることを目的として、従来のシステムに以下3.2節に示す改良を加えた。

### 3.2 ディスプレイ上への情報表示

システムからの返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3種類の手段によって実現した。

- 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)を表示。
- システムからの返答文が多く(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)情報を、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。

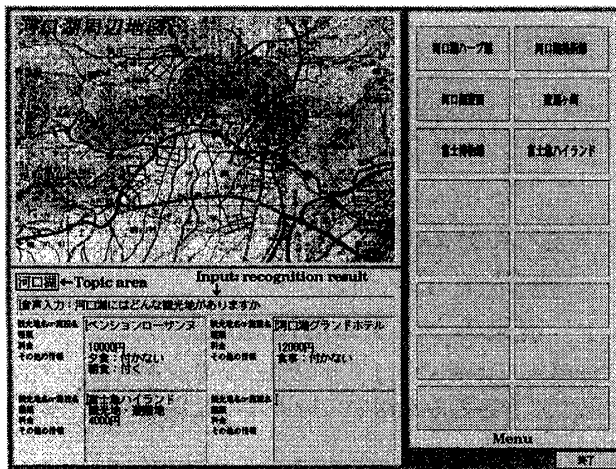


図 1: 画面表示の例

(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多くの内容を含んだ返答文の一部を聞き逃してしまった場合に、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したいと思うユーザを助けるため、ディスプレイ上に表示されたメニュー内の項目、若しくは、地図上の位置や地名等に触れながら(タッチ入力)、これを指示代名詞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た。指示代名詞&タッチスクリーン入力法によるシステムへの入力は例えば次のようになる。

ユーザ動作: メニュー内の項目「富士急ハイランド」にタッチする。

ユーザ発声: ここの入場料金はいくらかかりますか。

今回の実現は、入力音声認識部の認識結果が指示代名詞を含んでいるか否かを調べ、含んでいる場合に、指示代名詞とタッチ入力で決定している単語とを入れ換えるだけの処理しか行っていない(図2)ため、改良を要する。

## 4 おわりに

本研究では、従来音声のみをマンマシンインターフェイスとしていた音声対話システムに対し、対話の途中経過のディスプレイ上への表示、及び、タッチスクリーン入力法によって実現される、指示代名詞を含んだユーザ発話を受理する機能を実現した。

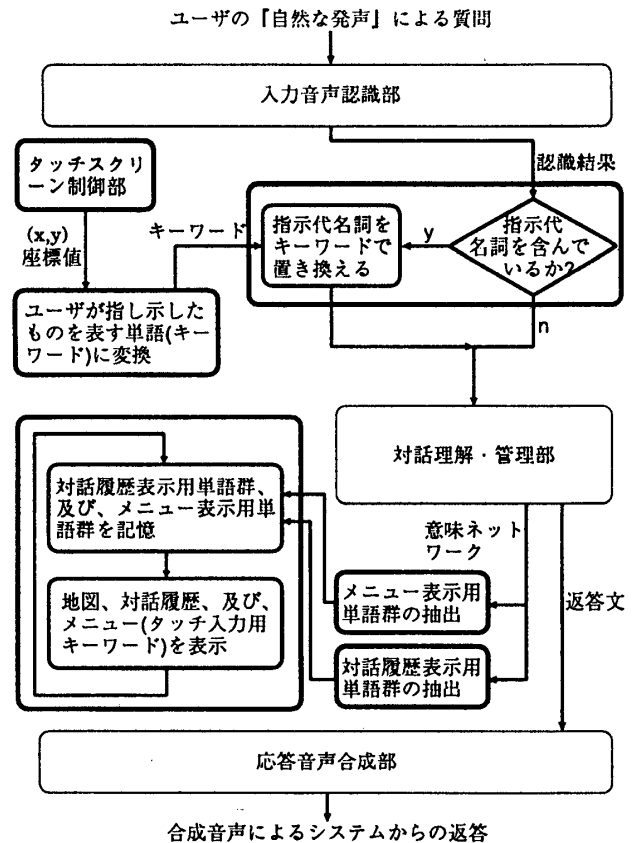


図 2: 改良後のシステムの概略図

今後は、マルチモーダルインターフェイス化後の「富士山観光案内日本語音声対話システム」の主観的、或は、客観的な評価を行っていく予定である。

## 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イス」, 電気学会論文誌, Vol.114-C, No.10, pp.1009-1017(1994)
- [3] 山本, 肥田野, 伊藤, 甲斐, 中川:「自然発話の意味理解と対話システム」, 情報処理学会, 音声言語処理研究会, 94-SLP-2-13(1994.7)
- [4] 伊藤, 肥田野, 山本, 中川:「ロバストな対話システム構築に関する一考察」, 情報処理学会, 音声言語情報処理研究会, 95-SLP-7-22(1995.7)