# マルチモーダルインターフェースと協調的応答を備えた観光案内対話システムの評価

伊藤敏彦 傳田 明弘 中川聖一

豊橋技術科学大学 情報工学系

〒441 豊橋市天伯町字雲雀ヶ丘 1-1

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」の開発において、ユーザにかかる負担や不安を軽減することを目的とし「タッチ入力、及び、指示詞を含んだユーザ発話の許可」や「対話の途中経過表示」などの機能を有するマルチモーダルインタフェースの実現、及び、システムからの質問やデータベース検索失敗時の代案提示などの協調的応答について検討している。

本稿では、「システムの使い勝手の良さ」が1) マルチモーダルインターフェイスの導入、2) 協調的応答生成の導入、によってどのように向上するのかに着目して行なった評価実験について述べる。

# Evaluation of a sightseeing guidance dialogue system based on spoken Japanese with multi-modal interface and cooperative responce

In our laboratory, a Japanese spoken dialogue system for Mt. Fuji sightseeing guidance has been developed. In the development, in order to reduce burdens imposed on a user, the following two functions to the conventional system were added; 1) a multi-modal interface where dialogue histories are shown on a display and additonal input methods are allowed such as touch screen input or spoken sentences including demonstratives. 2) a cooperative responce generater which gives users alternative proposals when system cannot retrieve an entry of databases directly satisfying usr's requirements.

Evaluation experiments are described where how the above improvement increses "convenience of the system" is focused upon.

## 1 はじめに

音声認識技術、及び、これを支援する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な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し、ユーザに使い勝手のよいマン-マシンインタフェースを提供することを目的として、「音声入出力」以外に「ポインティングジェスチャ入力」や「画像による出力」などの複数の入出力手段 ( モダリティ ) を相補的に統合し、同時あるいは逐次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なわれるようになってきている [1, 2, 3, 4, 5]。

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」[6, 7] 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なわれるために、ユーザに不安や負担を与えることがあった。そこで、「システムとの対話の途中経過表示」や「タッチ入力、及び、指示詞や指示代名詞 ( 以降は、これらをまとめて「指示語」と呼ぶ ) 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 [11] を試みている。

また同様に対話システムにおいて、システムがユーザと協調的に対話を進めていくことはユーザの負担を軽減するために重要である [12]。データベース検索における協調的応答生成に関しては質問の答以外に付加的な情報を与えたり、失敗した質問に対する理由や代案を提示するものが多い [13]。ユーザの質問文に検索に必要な情報が含まれていなかったり、検索結果の数が多い場合などはユーザへの質問を行なったり、ユーザの望む検索結果が得られなかった場合、それに代わる代案を提供する。このようなユーザへの協調的応答によってユーザにかかる負担や不安を軽減することも試みている [14]。

本稿では、同システムについて、「システムの使い勝手の良さ」、「マルチモーダルインタフェース

化の効果」、「協調的応答」に着目して行なった評価実験について述べる。

## 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」[6, 7] 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」、「多様な言い回し」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。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。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 [9]。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 ( もちろん、第 5 位の認識結果まで用いれば、この「対話理解・管理部」の性能は向上する [10]。 ) 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」を行う、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 [8]。

「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 システムの問題点と改良

### 3.1 従来のシステムの問題点

従来のシステム [7] では、マン-マシンインターフェイスとして音声のみを用いていることやシステムが不完全なために、以下のような問題点があった。

1. システムが Q&A システムに近いシステムであり、質問文以外の文をほとんど処理できない。
2. 使用できる文型に制限がある。
3. ユーザの提示した検索条件が少ない場合など、システムからの応答文が多くの内容を含んで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。
4. 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。
5. 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。
6. 音声認識部の処理時間が実時間の数倍程度で、自然な対話とは言い難い。

このように、発話できる文に制限があることや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、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。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、従来のシステムに次節に示す改良を加えた。

### 3.2 言語処理（理解）部の改良

これまでの言語処理（理解）部は、正しく意味表現に変換処理できる文型にかなりの制限があった。システムがもともと Q&A システムに近いシステムであったため、疑問文以外の文はほとんど正しく意味理解できなく、また副詞や形容詞を含んだ文の発話も許していなかった。さらに指示詞などの処理も完全に行なってないため、文を正しい意味表現に変換することは難しかった。

しかし、今回の言語理解部の改良により、疑問文以外の願望・依頼といった発話や副詞や形容詞を含んだ文も正しく意味理解できるようになった。さらに指示詞データベースの使用により、応答文や自分の発話に対する指示詞の使用もある程度正しく処理できるようになった。

### 3.3 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3種類の手段によって画面上に表示する。

- 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の写真などを表示。
- 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。

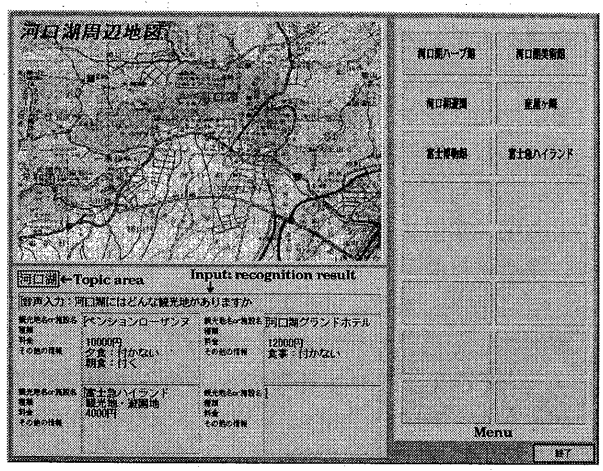


図 1: 画面表示の例

(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)

### 3.4 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た。指示語を含んだユーザの発話とタッチ入力を組み合わせた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。

### 3.5 協調的応答生成システム

問題点で述べたようにユーザの発話文から得られた検索条件が少ない場合は、知識データベースの検索に失敗したり、逆に大量の検索結果データを応答として出力したりする。

これはシステムがユーザの発話した一文だけから得られた検索条件で知識データベースを検索し、知識データベースから得られた情報を全てユーザに提示しようとするためである。さらにシステムは知識データベースを検索した結果、データが見つからなかった場合、知識データベースにデータがなかったことをユーザに示すだけで終る。そのため、あるタスクを達成しようとした場合に非常に対話数が多くなることがあった。

このような問題点を改良するためにユーザの対話の意図（焦点）を抽出しそれに沿った対話制御、検索条件が少ない場合のユーザへの質問、検索結果が見つからない場合の代案検索などができる協調的応答生成システムを構築した[15]。

今回、構築した応答生成システムの構成図を図2に示す。ユーザの発話した発話文は言語理解システムによって、図3のような等価な意味ネットワーク(意味表現)に変換され、最初に対話制御部に入力される。

対話制御部は対話の流れ、文脈情報の管理、必要な情報（条件）の質問などを行なっている。対話制御部は対話の流れを決定するために、ユーザの対話の意図(焦点)を抽出する意図解析部にユーザの発話意味ネットワークを送る。意図解析部は入力されたユーザの発話意味ネットワークからユーザの対話の意図(焦点)とユーザの提示した検索条件を抽出する。そして対話制御部は獲得したユーザの対話の意図(焦点)から対話の流れを決定し、これまでの発話から獲得された情報で現在の発話にも移行できる情報(文脈的な情報)があればそれを文脈情報として利用する。

次に知識データベースを検索するための問題解決器にはユーザの発話意味ネットワークと先ほど述べた文脈情報を入力する。問題解決器はそれらのデータをもとに知識データベースを検索する。

この時、問題解決器はユーザの発話意味ネットワークに含まれている検索条件と文脈情報を活用して検索するが、もし検索結果が得られなかった場合は検索条件の一部を同じ概念の検索条件に変更し検索をやり直す。そこで検索結果が得られたなら、代案としてその検索結果を対話制御部に送る。

対話制御部は獲得された検索結果を調べ、検索結果の数が多い場合には対話制御部はユーザへ検索結果選択のための質問を行なう。その時の質問はユーザの対話の意図 (焦点) に沿ったまだ獲得されていない検索条件に関して行なう。検索結果や代案のデータ数が適当であるなら検索された情報をユーザへ提示するために、対話制御部は応答文生成部へユーザの発話意味ネットワークと検索結果 (代案結果も含む) を送る。

応答文生成部では、入力された発話意味ネットワークと検索結果からどのような形で応答すれば良いかを決定し、それに従い応答文意味ネットワークを形成する。それから応答文意味ネットワークを通常の文字列に変換し、ユーザへの応答として音声合成で出力する。

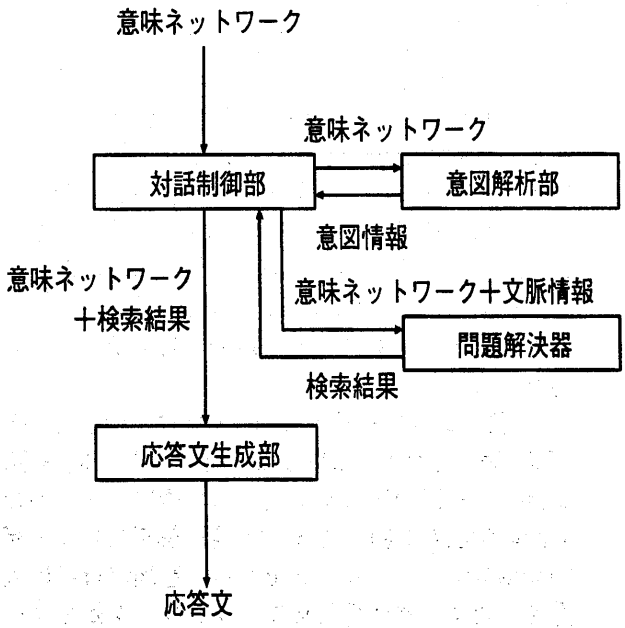


図 2: 応答生成システムの構成図

# 4 言語処理部改良に関する評価実験について

前節で述べたいくつかの改良点に関して、その効果を評価するために、実験を行なった。システムの評価実験ではマルチモーダルインターフェイス化による効果と言語処理部の改良（言語理解部と協調的応答生成システム）による効果をはっきり分けるために 2 種類の評価実験を行なった。ここでは言語理解部と協調的応答システムの部分に焦点をあてた評価実験について述べる。次節でマルチモーダルインターフェイス化に関する評価実験を述べる。

発話文：
河口湖でテニスのできるホテルはありますか。
‖
発話意味ネットワーク：

```
(ある (FORM YN-Q)(TARGET NIL)
      (NEGATION NIL)
      (OBJ (ホテル ((AT-LOC)
                    (できる (OBJ (テニス))))))
      (AT-LOC (河口湖)))
```

図 3: 発話意味ネットワーク例

## 4.1 実験の形式・条件

実験は、被験者らに「1 泊 2 日の富士山周辺への研究室旅行」を、システムとの対話で得られる情報をもとに計画してもらうという形式で行なった。決定してもらう内容は、1、2 日目の目的地とそこでのプラン、及び、宿泊先の施設の所在、種類、料金、宿泊施設名の計 8 項目である。被験者は、音声対話システムに関して全く知識のない本大学生（学部 2〜3 年）6 人であり、特にシステムの使用法は教示せず、音声認識モデルの話者適応化用に発声してもらったタスクに関する 50 文を発声例として自己学習するだけである。

実験では以下の 2 種類のシステムで計画を立ててもらった。

1. 従来の対話システム（system1）
2. 言語処理部改良後の対話システム（system2）

音声認識システムはどちらも従来のものを使用している (2 節参照)。実験に使用した認識条件はケプストラム、回帰係数使用、ビーム幅 5 である（認識を高速に行なうためにビーム幅を狭くし、また音声の継続時間分布制御も使用していない）。また HMM には研究室内の雑音環境に適応化した不特定話者モデルを、実験前に被験者に発話してもらった富士山観光案内に関した適応化文 50 文で話者適応化し、使用している。音声認識用文法は単語数 275、パープレキシティ約 100、である。自然な発話を受理するように文法にロバスト性をもたしたためパープレキシティは非常に大きくなった。しかしながら、音声認識処理に有する時間は発話時間を含め実時間の 1.5〜2 倍程度（HP-C160 使用) で、比較的時間的にはスムーズな対話が続行できるようになった。

実験の手順としては、最初に実験の内容と決定した項目を記入してもらう紙を渡し、熟読してもらった。その後、system1 と system2 のそれぞれとの対話（最初に対話してもらうシステムは被験者毎に変更）で旅行計画を 2 案立ててもらった。実験終了後、使用感や応答に関するアンケートに答えてもらった。

## 4.2 実験の結果

評価実験における被験者のデータを表 1に示す。項目達成は 8 項目のうちどれだけ決定できたかである。発話数は実験終了までに発話した文数である。発話数の項目の括弧内の数字は system1 と system2 を使用した順番である。まず、この 6 人の項目達成率 ( 必要な 8 項目を決定できた割合 ) は、system1 で平均 94%、system2 で平均 100%であった。その理由として system1 の対話システムでは宿泊施設の料金に関する情報を聞き出すことができなかった人が多かった。全発話数に関してはどちらのシステムも大きな差はでなかった。system1 と system2 での発話数に大きな差がでなかった理由として、第一には system2 の場合、システムの質問（9 回）に応答する発話数が含まれていることがある。第二にはあまりにタスクが簡単すぎたという理由も考えられる。もう少しタスクが複雑なものであればもっとはっきりと発話数に差がでたのではないかと思われる。次に個々の被験者の発話数を見ていくと慣れのためと思われるが、2 番目に行なったシステムの方が発話数は減少する傾向がある。しかし一項目当たりに必要な発話数を見ると sytem1 で 2.3、system2 で 2.1 であり、改良による労力軽減の効果が現れているとは言える。

表 1: 評価用データ

| | system1 | | system2 | |
|---|---|---|---|---|
| 被験者 | 項目達成 | 発話数 | 項目達成 | 発話数 |
| A | 8/8 | 12(1) | 8/8 | 12(2) |
| B | 7/8 | 29(1) | 8/8 | 12(2) |
| C | 8/8 | 8(1) | 8/8 | 11(2) |
| D | 8/8 | 6(2) | 8/8 | 13(1) |
| E | 7/8 | 13(2) | 8/8 | 15(1) |
| F | 7/8 | 35(2) | 8/8 | 38(1) |
| 合計 | 45/48 | 103 | 48/48 | 101 |

次に表 2にシステムの評価結果を示す。正解音声認識率は被験者が発話した文が正確に正しく認識された割合を示している。認識文理解率は、(終) 助詞落ち・誤りなど人間が見れば誤りと分かるものを許した場合の音声認識率である。正解意味理解率は被験者の発話が言語理解部で正しい意味表現に変換できた割合である。情報提供率はユーザの発話に対して正しい情報を提供した割合である。システム質問率はシステムがユーザに情報選択するための質問をした割合、代案提供率はユーザに検索結果の代わりに代案を示した割合である。正解応答率は情報提供率、システム質問率、代案提供率の他に意味表現作成に失敗した割合（システムは発話理解に失敗したことを応答する）を加えたものである。

表 2: 評価結果

| 評価 | system1 | system2 |
|---|---|---|
| 全発話数 | 103 文 | 101 文 |
| 正解音声認識率 | 32 文 (31.0%) | 21 文 (20.8%) |
| 認識文理解率 | 83 文 (80.6%) | 56 文 (55.4%) |
| 正解意味理解率 | 63 文 (61.2%) | 56 文 (55.4%) |
| 正解応答率 | 63 文 (61.2%) | 81 文 (80.2%) |
| 情報提供率 | 49 文 (47.6%) | 43 文 (42.6%) |
| システム質問率 | 0 文 (0%) | 9 文 (9.0%) |
| 代案提供率 | 0 文 (0%) | 3 文 (3.0%) |

正解音声認識率、認識文理解率を見ると同じ音声認識システムを使用しているにも関わらず大きな差がでている。これは sytem1 と system2 では発話された文の種類がかなり異なっているためである。前節でも述べたように system1 と system2 の処理できる文の種類に大きな差がある。そのため system1 で最初にある文を発話し、それが正しく理解されないことが分かるとそれ以後、被験者は system1 ではそれに似た文を発話しない。そのため比較的認識されやすい文ばかりが発話されたためにこのような結果になったと思われる。認識文理解率と正解意味理解率の差が少ないほど言語理解部の能力が人間なみであると言える。今回の改良によってかなり性能が向上していることが分かる。しかし、システムの質問に対するユーザの応答が誤認識されるとシステムは誤った情報を提供したり、発話の誤認識によって文脈知識データが誤って更新されたりすることがあった。これらの失敗に対する対応（ユーザへの確認等）が必要であると考えられる。

表 3に音声認識率 100%として被験者の発話を

書き起こしたものを入力（テキスト入力）とした場合の評価結果を示す。発話された文の種類の豊富さにも関わらず、正解意味理解率、正解応答率共にsystem1よりsystem2の方が大きく上回っている。このことからも表2の正解意味理解率の結果が、音声認識の難しさの影響を大きく受けていることが分かる（人間が見ても理解できない認識文は言語処理部でも難しい）。次にシステムがユーザにタスク達成のために有益な何らかの情報を提供する割合がsystem1の57.3%からsystem2の68.4%（63.4%+5.0%）と大きく上昇している。この割合の上昇はタスク達成に費やす労力を大きく軽減すると思われる。

表3: 評価結果（入力：書き起こしデータ）

| 評価 | system1 | system2 |
|---|---|---|
| 全発話数 | 103文 | 101文 |
| 正解意味理解率 | 73文(70.9%) | 90文(89.1%) |
| 正解応答率 | 69文(67.0%) | 87文(86.1%) |
| 情報提供率 | 59文(57.3%) | 64文(63.4%) |
| システム質問率 | 0文(0%) | 12文(12.0%) |
| 代案提供率 | 0文(0%) | 5文(5.0%) |

### 4.3 使用感に関するアンケート

実験終了後、言語処理部に関するアンケートを被験者に書いてもらい使用感などを調査した。計画の立てやすさ、システムの便利さに関しては、質問や代案を提示したり、処理できる文の種類が豊富なsystem2の評価が高い。応答の聞き易さ、応答の自然さ、次発話のし易さといった項目ではsystem2が少し良い程度であった。評価実験では質問や代案提示の回数がかなり少ないため、あまり差がでなかったと思われる。

次にそれぞれのシステムでうまくいった対話例を書き起こして示し、システムの便利さ、システムの自然さ、次発話のし易さの項目を質問した場合はどれもsystem2の評価がかなり高かった。

次に対話制御に関するアンケートでは「検索結果を取捨するためのシステムからの質問は必要か」という質問に半分が必要、半分が不必要（検索結果を全部提示してほしい）と答えた。代案に関しては全員があると便利であると答え、計画がうまくいかない時のシステムからの提案に関しても全員が望んでいた。またシステム主導型のような必要なことを全部聞いてくれるシステムが便利であるという意見もあった[17]。

これらのアンケートの結果からも、今回の改良による労力軽減の効果はあったと考えられる。

## 5 マルチモーダル音声対話システムの評価について

本節では、マルチモーダルインタフェース化を行なった「富士山観光案内日本語音声対話システム」について、「システムの使い勝手の良さ」や「マルチモーダルインタフェース化の効果」に着目して行なった評価実験について述べる。今回の実験に対する予備実験として行なった評価実験[16]では、満足のいく「ユーザ発話認識率」（後述）が得られていない。原因としては指示語を含んだ文に対する音声認識用文法の不備や、前述した言語処理部の不完全さなどの問題点が浮き彫りになったためである。今回の実験前に、音声認識用文法の改良を行ない、言語処理部には改良後の言語処理部を使用した。

実験の形式・条件は4節の言語処理部の評価実験の形式・条件を参照されたい。被験者は、音声対話システムに関して全く知識のない本大学生（学部2〜3年）10人である。被験者10人の項目達成率は、約79.2%であった。

### 5.1 対話の形式

評価実験では、以下の3種類の形式でシステムとの対話を行なってもらい、それにより旅行計画を3案立ててもらう。括弧内の記述は、それぞれの表における各対話形式に対応している。

1. 音声のみの入力、及び、音声のみの出力による対話（音声入出力）
2. 出力は音声出力に加え、対話の途中経過をディスプレイ上への画像出力で与えるマルチモーダル出力とし、入力は音声のみで行なう対話（音声入力・マルチ出力）
3. 入力は音声のみ、または音声とタッチスクリーンを用いた入力の併用で実現されるマルチモーダル入力、出力は音声及びディスプレイ上への途中経過画像によるマルチモーダル出力で行なわれる対話（マルチ入出力）

今回の実験では、被験者A,D,G,Jには対話形式1→2→3、被験者B,E,Hには、対話形式2→3→1、被験者C,F,Iには、対話形式3→1→2の順で対話を行なってもらった。

### 5.2 評価結果

表4に「システムのユーザ発話理解率と課題達成のために費やされた発話数」を示す。

表 4: ユーザ発話理解率と課題達成のために費やされた発話の数

上段：ユーザ発話認識率 [%] 中段：ユーザ発話理解率 [%] 下段：ユーザ発話数（括弧内 タッチ入力を用いた発話数）
（*印は 1 番目に行なった対話形式）

| 対話形式 使用順 | 被験者 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J |
| 音声入出力 | 45.8*<br>45.8*<br>24* | | | 45.5*<br>45.5*<br>33* | | | 38.9*<br>38.9*<br>18* | | | 27.5*<br>15.9*<br>69* |
| 音声入力・マルチ出力 | 62.5<br>62.5<br>24 | 55.6*<br>44.4*<br>9* | | 75.0<br>75.0<br>12 | 43.8*<br>34.4*<br>32* | | 34.3<br>40.0<br>35 | 40.0*<br>35.0*<br>20* | | 37.3<br>29.4<br>51 |
| マルチ入出力 | 77.7<br>55.6<br>9 (0) | 80.0<br>80.0<br>5 (0) | 25.0*<br>37.5*<br>16 (8)* | 54.5<br>40.9<br>22 (10) | 46.9<br>40.6<br>32 (9) | 46.3*<br>46.3*<br>41 (5)* | 47.8<br>47.8<br>23 (2) | 40.0<br>40.0<br>15 (1) | 35.7*<br>33.3*<br>42 (2)* | 54.8<br>41.9<br>31 (10) |
| 音声入出力 | | 66.7<br>50.0<br>6 | 34.5<br>27.6<br>29 | | 40.9<br>40.9<br>22 | 70.6<br>70.6<br>17 | | 53.8<br>53.8<br>13 | 37.0<br>37.0<br>27 | |
| 音声入力・マルチ出力 | | | 10.0<br>10.0<br>10 | | | 66.7<br>75.0<br>12 | | | 56.0<br>44.0<br>25 | |

「システムのユーザ発話理解率」は、対話システムがユーザの発話を正しく理解できた割合である。本実験においては、ユーザ発話から正しい意味ネットワークに変換された数とユーザの全発話数によって、以下のように定義する。

$$\text{ユーザ発話理解率} = \frac{\text{正しい意味ネットワークに変換できた数}}{\text{ユーザ全発話数}}$$

表中の＊の付いた（1 番目に実験を行なった対話形式）結果は、2 番目以降の結果とシステムへの不慣れさによる差があるため評価結果としては除外する。

表 4 より、多くの被験者が音声のみで行なわれる対話よりも、マルチモーダル化したシステムにおいて行なわれる対話でより高い「音声認識率」、「ユーザ発話理解率」を示している。これは音声のみのシステムでのユーザ発話が、発話内容を考えながら（システムの応答を思い出しながら）発話するのに対して、マルチモーダル化したシステムでは発話するための情報や応答が表示されているためか、発話前に発話内容がかなり決定しているらしくかなり定型的な発話になっていた。

また、「課題達成のために費やされた発話数」を見ると、被験者 A,E,F,H,I,J は、「富士山観光案内システム」に対して、徐々に慣れていく様子が見受けられるが、マルチモーダル対話を行なった後に、特に発話数の減少が大きい。

これらのことから、音声対話システムの知識を持たないユーザに対し、マルチモーダルインタフェースは、音声対話システムに理解されやすい対話と、対話システムへの素早い慣れを実現できる可能性と言える。

## 5.3 アンケートによる考察

「ディスプレイ上への途中経過表示」について「地図表示」については、役に立ったと答えた被験者が 4 人、役に立たなかったと答えた被験者が 4 人という結果となった。役に立たなかったと答えた被験者は、地図上の字が読みにくいことや、地図上に表示されている地名や観光地名が、まだ完全に言語処理部で対応しきれていない（辞書の未登録など）ことを指摘した。

「メニュー表示」については、多くの被験者が役に立ったという返答をしてくれた。中には、「これがないと何を聞いてよいのか、さっぱりわからなかった」という意見もあり、計算機とコミュニケートすることになれていないユーザの対話を手助けする役割を果たしている様子が見受けられた。

「対話履歴表示」については、これといった反応は見られず、可もなく不可もなくという感じであった。中には、「表示される程度の情報（観光地・施設名）は耳で聞き取れた」として不必要とする意見もあった。

「タッチ入力」について半数の被験者が「タッチ入力は使いにくかった」と答えた。その理由として、「タッチ入力の精度が悪い」と答えた被験者が 3 人いた。彼らの対話におけるタッチ入力の語動作の内訳を調べてみると、

意図したものが指し示せなかった：9 例
指示語を含んだ文の音声認識誤り：9 例
発話に対して応答が正しくない：1 例
データ転送時の誤動作：1 例

となっていた。一方で、「どのような発話（タッチパネル使用時）をすればいいのか思いつかない (2人)」「指示語を含んだ文のバリエーションが少ない (1人)」といった理由で、タッチ入力を「使えなかった」または「使わなかった」被験者もいた。ただ、どの被験者も、タッチ入力の性能が向上すれば、「大変役に立つと思う」「そこそこ役に立つと思う」と答えているので、今後、ハードウエア及びソフトウエアの両方の面から、タッチ入力インタフェースを改良していきたいと考えている。

「どの入力インタフェースが使いやすいか」上述したように、ほとんどの被験者が「タッチ入力」を使いにくいとしているため、「音声入力・マルチ出力」の入出力インタフェースが使いやすいと回答が大勢を占めた。

そのほかの意見としては、システム全体に対するものとして、「『音声入出力』『画像出力』『タッチ入力』のそれぞれが一つにまとまっていない、かみあっていない気がする」という意見もあった。

以上のようなアンケートからもマルチモーダルインターフェイスの対話システムの有用性は証明することができた。今後はタッチ入力の改良を中心としたシステム全体の整備を行なっていく必要がある。

## 6 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ー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「タッチ入力と、指示語を含んだユーザ発話を組み合わせた入力」を実現したマルチモーダル化の改良、さらにユーザに協調的に応答する言語処理部の改良を行なった。またシステムの評価実験においては、言語処理部の改良によって、タスク達成への労力の軽減が見られた。さらに対話システムへのマルチモーダルインターフェイス化の有用性を示すことができた。

今後は現在のシステムに対して、評価実験によって明らかになったシステム全体の不備な点を改良していく予定である。

## 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] 安藤, 北原, 畑岡：「インテリアデザイン支援システムを対象としたマルチモーダルインタフェースの評価」, 電子情報通信学会論文誌, VOL.J77-D-II, No.8, pp.1465-1474(1994).

[3] 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)

[4] 神尾, 松浦, 正井, 新田:「マルチモーダル対話システム MultiksDial」, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1429-1437, (1994).

[5] 伊藤, 古川, 中沢, 木山, 張, 岡：「複数ユーザによる音声とジェスチャのマルチモーダルインタフェースシステム：Real-time GSI の一評価実験」情報処理学会, 音声言語情報処理研究会報告, 96-SLP-10, pp.3-8 (1996)

[6] 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)

[7] 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, (1996.4).

[8] 山本、小林、中川：「音声対話文における助詞落ち・倒置の分析と解析手法」、情報処理学会論文誌 Vol.33, No.11, pp.1322-1330(1992).

[9] 甲斐、間宮、中川：「自然発話の認識・理解のための解析・照合手法の比較」、情報処理学会、音声言語情報処理研究会報告、94-SLP-2-12(1994.7)

[10] 肥田野 勝, 中川 聖一: 「音声対話システムにおける N-best 文認識結果の一利用法」, 情報処理学会第 52 回全国大会 (2), 4D-2, pp.165-166, 1996.

[11] 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

[12] Kaplan,S.J.:Cooperative Responces from a Portable Natural Language Database Query System,Brady, M. and Berwick R.C. (eds.), Computational Models of Discourse, pp.167-208, MIT Press(1983).

[13] Webber,B. :Question Answering, Shapiro,S.C.(ed.), Encyclopedia of artifical intelligence, pp.814-822, New york: Wiley(1987)

[14] 伊藤, 中川：「音声対話システムにおける協調的応答」, 情報処理学会, 音声言語情報処理研究会報告, 96-SLP-10, pp.105-110 (1996)

[15] 伊藤, 中川: 「音声対話システムにおける焦点の抽出と協調的応答」, 情報処理学会第 53 回全国大会 (2), 7N-4, pp.353-354, 1996.

[16] 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」 , 人工知能学会全国大会 (第10回) 論文集, pp.431-434, 1996

[17] 中川聖一 , 山本誠治 : 「音声対話システムの構成法とユーザ発話の関係」 , 電子情報通信学会論文誌 , Vol.J79-D-II, No.12, 1996