

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実験

傳田 明弘 伊藤敏彦 小暮 悟 中川聖一

E-mail address : {akihiro, itoh, kogure,nakagawa}@slp.tutics.tut.ac.jp

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

近年、音声対話システムの対話を支援し、使い勝手の良いインタフェースをユーザに提供することを目的としたマルチモーダルインタフェースの研究が盛んになってきている。

音声対話システムの入出力のインタフェースをマルチモーダル化することは、より自然で内容豊かなマンマシン対話を実現するために有用なことである、という考えの元に、本研究室では、富士山周辺の観光案内をタスクとする「富士山観光案内日本語音声対話システム」に「タッチ入力」及び「対話の途中経過の画面表示」の機能を付加した、マルチモーダルインタフェース化の改良を行なっている。

我々は、今回、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を使用して、あるタスクを達成してもらう被験者実験を行なった。本稿では、その結果について述べる。

Evaluation Experiment of A Sightseeing Guidance Spoken Dialogue System with Multi-Modal Interface

Akihiro DENDA, Toshihiko ITOH, Satoru KOGURE and Seiichi NAKAGAWA

E-mail address : {akihiro, itoh, kogure,nakagawa}@slp.tutics.tut.ac.jp

Department of Information and Computer Sciences

Toyohashi University of Technology

Tenpaku-cho, Toyohashi-shi, Aichi-ken, 441, Japan

Recent improvements of speech recognition and natural language processing enable dialogue systems to deal with spontaneous speech. With the aim of supporting these systems, multi-modal man-machine interface has been introduced to the system widely. In addition to increasing the total performance of the systems, the multi-modal interface is expected to make the dialogues between a user and the system more natural and abundant in content.

In our laboratory, we have developed the multi-modal interface with speech input/output, graphical output and touch input for our spoken dialogue system, "Mt. Fuji Sightseeing Guidance System by Spoken Japanese".

And we had made evaluation experiments, where we had investigated the usefulness of our multi-modal interface.

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 [8] 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 [9]。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 [10]。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 [11]。

本稿では、同システムのマルチモーダルインタフェースについて、「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」といった点に着目して行なった評価実験について述べる。

2 富士山観光案内システム

「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」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。

「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つの部分に、以下で述べる「マルチモーダルインタフェース」を付加した構成になっている [12]。

2.1 マルチモーダルインタフェース

観光案内のような検索システムのインタフェースが、音声のみで実現されている場合には、従来のシステムでは、様々な問題点が生じてくる [12] ため、場合によってはユーザに不安や負担を与えかねない。そこで、問題点の一つの解決策として本システムでは、以下に示すようなマルチモーダルインタフェースを採用している。

2.1.1 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。

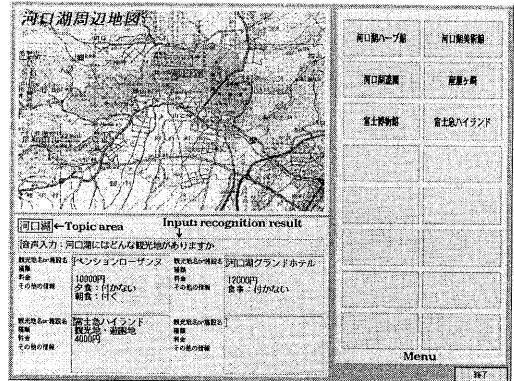


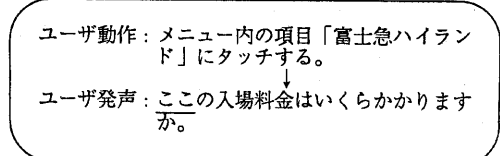
図 1: 画面表示の例
(左上: 地図, 左中央: 現在のトピックと入力音声の認識結果, 右: 応答文のメニュー, 左下: 対話履歴)

- 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。

2.1.2 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「こ」系列の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。「こ」系列の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、例えば次のようになる。



3 マルチモーダル音声対話システムの改良点

予備実験として行なった評価実験 [12] では、被験者らは本対話システムの知識を有する者（未経験者）であったが、満足いく「ユーザ発話認識率」が得られない、指示語を含んだ文に対する文法規則やユーザ発話の認識結果から意味ネットワークに変換する生成規則が不十分である、といった問題点が浮き彫りにされたため、今回の実験の前に、以上の問題点を解消する改良をまず行なった。

3.1 「入力音声認識部」の改良

以前行なった予備評価実験 [12] では、処理の高速化のために継続時間制御やケプストラムの回帰係数を使用しない条件で認識を行っていたが、現在のシステムの「入力音声認識部」では、回帰係数を使用している。継続時間制御は、使用していない場合に比べ、認識時間が3~4倍余計にかかってしまい、リアルタイム応答が不可能になってしまうため、今回も使用していない。また、リアルタイム応答を実現するため、認識用の計算機を、HP Model J200 から、HP Model C-160 に変更することによって、認識の高速化をはかり、認識時間はリアルタイムの2~3倍程度になった。

認識に使用している CFG についても、前回の予備評価実験では、文法記述の不十分さがあったため、これまで扱えなかった指示語を含む文や格助詞相当句（…について、…に関して）を含む文などを扱えるよう、新たな文法記述を追加し、併せて、単語辞書にいくつかの単語も追加した。これにより、パーレキシティは、従来の74から103に、単語辞書の語彙数は、245から292に増加している。

これらに加え、話者適応化の際に基本となる HMM 音声モデルにも改良を加えている。従来の基本 HMM 音声モデルは、以下のような手順で作成したものを使用していた。まず、ATR 研究用日本語音声データベースの6名分の発声（503文/人）を用いて基本モデルを作成する（発話者は男性。以後、話者はすべて男性である）。次に、日本音響学会研究用連続音声データベースの30名分の発声（約150文/人）を用いて基本モデルに対し追加学習を行なった。今回は、この基本 HMM モデルに対し、本研究室の計算機室の環境（ノイズあり）で収録した10名分の発声（50文/人）によって、環境適応化の追加学習を行なったモデルを、新たな適応化用基本 HMM 音声モデルとして使用している。

3.2 「対話理解・管理部」の改良

これまでの、「対話理解管理部」では、形容詞や副詞を含んだ文、「そ」系列の指示語を含んだ文（& 文脈による指示語の補完）を正しく解析することができなかった。そこで、「対話理解・管理部」において形態素解析システム「JUMAN」の辞書に、「富

士山観光案内システム」で使用されると予想される形容詞・副詞（「安い」「近い」「一番」「最も」など）を追加し、形容詞・副詞を含んだ文を、意味ネットワークに変換するための新たな変換規則を追加する改良を行なった。このことにより、次のような文を正しく意味理解できるようになった。

- ・ 河口湖で一番安いホテルを教えてください。
- ・ 富岳の風穴は涼しいですか。

但し、現在のシステムでは、これらの文が正しく意味理解されても、ユーザの要求を満たすような知識データベース検索が実現されていないため、正しい応答を返すことができない。例えば、上記のユーザ発話例に対しては、システムから次のような応答が返される。

- ・ レイグランドホテル、河口湖ホテル、河口湖グランドホテル、ハイランドホテルが河口湖にはあります。（※ 河口湖にある全てのホテルを返答する）
- ・ データベースにないのでお答えできません。

また、「そ」系列の指示語に該当する指示対象を、文脈の情報から補完する機能を新たに付加することも行なった。

4 評価実験

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。

4.1 実験の方法

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元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。なお、被験者らには、旅行計画を行なう際、どうしても必要な情報が対話で得られない場合には、その項目は決定しなくてもいい、という指示をあらかじめ与えてある。被験者は、音声対話システムに関する知識を持たない大学学部生10人である。被験者らには、実験の前に「富士山観光案内日本語音声対話システム」で実際に扱える入力文で構成された適応化文50文によって、認識に用いる HMM の適応化を行ってもらう。この、適応化文50文を読み上げによって、被験者らに本システムのタスクやシステムが扱える発話のおおよその範囲を知ってもらった。

4.2 対話の形式

話者適応化後、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、それにより計画を3案立ててもらい、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ている。各被験者の1番目の対話は、システム使用の不慣れのため評価から除外する。

実際の対話例

「マルチ入出力」の対話形式で実現される実際の対話の例を図2に示す。

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、富士山にはどんな湖がありますか。
SYS: 河口湖や山中湖、精進湖や本栖湖、西湖が富士山にはあります。
(富士五湖の各名称がメニューとして表示される)
USR: ここでテニスはできますか。
(メニュー内の「本栖湖」にタッチ)
SYS: テニスが本栖湖でできます。
USR: 本栖湖にテニスコートあるんですか。
SYS: 本栖湖テニスコートが本栖湖にはあります。
USR: 本栖湖テニスコートはいくらですか。
SYS: 本栖湖テニスコートは8000円かかります。
(本栖湖テニスコートの写真が表示される)
USR: 河口湖には何がありますか。
SYS: 設備や施設、見所や遊園地、博物館が河口湖にはあります。
USR: ど、どんな遊園地があるんですか、えーと、河口湖には。
SYS: 富士急ハイランドが河口湖にはあります。
USR: その入場料金はいくらですか。
SYS: 富士急ハイランドの入場料は4000円かかります。
(富士急ハイランドの写真が表示される)

図2: 対話例 (マルチ入出力)

4.3 評価結果

被験者10人のタスクの平均項目達成率(項目を決定できた割合)は、約80%であった。対話形

式間のタスクの達成率を比較してみると、「音声入出力」対話では、7人中4人がタスクを達成(8項目全てを決定)でき(7項目以上は5人)、「音声入力・マルチ出力」対話及び「マルチ入出力」対話では、7人中1人(7項目以上は両対話ともに3人)しかタスクを達成することができなかった。これは、音声のみの対話に比べ、マルチモーダル対話でシステムから得られる豊富な情報によって、次発話が文法の制限を越えたり知識データベースの範囲外の内容を尋ねるものになってしまうためであると考えられる。また、宿泊施設の料金の項目を決定できていない被験者が比較的多かった。この理由は、「宿泊施設名の項目が決定できないと、この項目は決定できないこと」、「特に2人の被験者は料金を一度だけ聞いて情報が得られなかったため、あきらめてしまった」ことによる。

「入力音声認識部」の文認識率(助詞誤りは無視)は、約50%で、そのうちの約90%、全体の約46%が「対話理解・管理部」における解析で、正しい意味ネットワークに変換できた。また、音声認識用の文法による発話の受理率は、約72%である。これに対し「入力音声認識部」の文認識率を100%と仮定した、被験者の発話書き起こしデータ(間投詞や言い直し部分は削除)をタイプ入力した場合の文理解率は、約76%であった。実音声の場合に比べて、タイプ入力の際の文理解率が芳しくないが、理解できなかった文を調べてみると、そのうちの2/3は、「対話理解・管理部」の辞書単語や意味ネットワーク生成規則の不足が原因であった。このことは、「対話理解・管理部」へのデータ登録がまだ必要であることを意味している。

表1にシステムの「ユーザ発話認識率」「ユーザ発話理解率」及び「1発話に費やした時間(対話に費やした時間÷発話数)」を示す。「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発話数によって、以下のよう

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

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1より、「ユーザ発話認識率」「ユーザ発話理解率」を対話形式「音声入出力」と「音声入力・マルチ出力」「マルチ入出力」とで比較してみると、それぞれの対話形式間に優位性は見られない。このことについての理由は幾つか考えられるが、その一つとして、音声のみの対話におけるユーザ発話とマルチモーダル対話におけるユーザ発話の傾向が異なっているため、ということがあげられる。

マルチモーダル対話では、システムから、システムの応答音声に地図やメニューといった画像情報を統合した形で情報提供がなされるため、その豊富

表 1: ユーザ発話認識率、理解率及び1発話に費やした時間

1 段目: ユーザ発話認識率 [%] 2 段目: ユーザ発話理解率 [%] 3 段目: 1 発話に費やした時間 (括弧内: その対話形式での総発話数)
 (*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* 45.8* 28'0"*(24)*			45.5* 45.5* 51'0"*(33)*			38.9* 38.9* 35'2"*(18)*				27.5* 15.9* 45'0"*(69)*
音声入力・マルチ出力	62.5 62.5 29'5" (24)	55.6* 44.4* 72'0"*(9)*		75.0 75.0 37'4" (12)	43.8* 34.4* 33'7"*(32)*		34.3 40.0 34'4" (35)	40.0* 35.0* 72'2"*(20)*			37.3 29.4 37'8" (51)
マルチ入出力	77.7 55.6 51'1" (9)	80.0 80.0 57'2" (5)	25.0* 37.5* 45'1"*(16)*	54.5 40.9 35'0" (22)	46.9 40.6 34'5" (32)	46.3* 46.3* 72'7"*(41)*	47.8 47.8 34'9" (23)	40.0 40.0 57'9" (15)	35.7* 33.3* 51'9"*(42)*		54.8 41.9 38'4" (31)
音声入出力		66.7 50.0 51'7" (6)	34.5 27.6 35'2" (29)		40.9 40.9 26'7" (22)	70.6 70.6 77'6" (17)		53.8 53.8 54'9" (13)			37.0 37.0 47'6" (27)
音声入力・マルチ出力			10.0 10.0 35'5" (10)			66.7 75.0 63'8" (12)					56.0 44.0 50'0" (25)

な情報によって、ユーザに文法外発話や知識データベースの範囲外の発話を促してしまうことがある。また、被験者によっては、タッチ入力の際に発話する「こ」系列の指示語が「これ」→「どれ」のような認識誤りをしやすい場合があり、このような被験者がタッチ入力を頻発する対話での「ユーザ発話理解率」は思わしくない場合が多い。これに対し、音声のみの対話では、システムから与えられる情報が、システムの応答発話に限られてしまうため、発話はシステムによって緩いながらも制御を受け、文法外やデータベースの範囲外の発話は比較的少ない。よって、マルチモーダル対話で、音声のみの場合と同じような対話をするユーザ (B, F) は、対話形式間に顕著な差は生じず、マルチモーダル対話で音声のみの対話と傾向の異なった対話をするユーザ (E, H, I) は、その対話の進め方によって、対話形式間で「ユーザ発話認識率」が異なってくることになる。

また、「1発話に費やした時間」を見ると、多くの被験者が、「音声のみの対話」よりも「マルチモーダル対話」で、各発話に時間をかけて行っている様子が見受けられる。これは、マルチモーダル化したシステムのディスプレイ表示から得られる情報を十分に吟味して次発話を決定しているためだと思われる。実際に被験者の対話の様子を観察してみると、画面を見ながら考え込んだり、地図上のどの位置に触れようかと思案しながら指を移動させたりする様子が何人かの被験者から観察できた。

これに合わせて、各形式間の発話内容を見てみると、マルチモーダル対話の際に、音声のみの対話では現れなかった発話が見られ、対話の内容が豊かになっている様子が見受けられる。例えば、ある被験者は、河口湖の地図で湖の上に引かれている線を見て、「遊覧船に乗りたいたいです」と発話した。

音声のみの対話で、「河口湖で遊覧船に乗れる」とユーザに返答することはないので、これはユーザが地図から得た情報を元に対話を行なって、対話のバリエーションが広がった一例と言える。また、ある観光地への行き方を尋ねた被験者もいた。地図上に、その観光地への道のりが表示されることを期待しての発話であろうと思われる。現在のシステムでは、「遊覧船」や「観光地への行き方」はサポートしていないため、有効な応答を返すことはできないが、知識データベースの整備によってこれらがサポートできれば、画像情報から対話内容を豊かにできるものと思われる。

このように、マルチモーダルインタフェースには、対話の内容を豊かにすることができる反面、システム側のより高度な処理技術 (データベースの充実を含む) が要求される。

4.4 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。つまり、地図から得られた付加的な情報に、対話システムが対応していないために不快感を感じたためであろう。しかし、言語処理部と地図の情報がきちんと対応づけば、地図という画像の情報で提供できる情報は、音声のみの対話で提供できる情報よりはるかに多く伝えることができ、さらにこれは対話の内容を豊富にすることもできることを示唆している。

「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た。中には、「これがないと何を聞いてよいか、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ションすることになっ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」「そこそ役に立つと思う」と答えているので、今後、ハードウェア及びソフトウェアの両方の面から、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いう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フェースの対話システムの有用性を示すことができた。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っていく必要がある。

5 おわりに

本研究では、従来音声のみをマンマシンインタ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を行った。そして、そのシステムを使用して、被験者によるインタフェース及びシステム全体の評価を

行なった。しかし、「ユーザ発話認識率」「ユーザ発話理解率」「対話に費やした発話数」「1発話に費やした時間」といった定量的な結果のデータから、マルチモーダルインタフェースの有用性を十分に示せたとは言えない。今後は、現在のシステムに対し、評価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・調整していくとともに、マルチモーダルインタフェースの有用性を示す客観的かつ定量的な評価尺度を模索していくつもりである。一方、アンケートによる主観的な評価では、マルチモーダルインタフェースの有用性を十分に示すことができた。

今後は、エージェントを導入するなど、計算機と対話を行なうユーザに、心理的な面からも「使い勝手が良い」と言える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、改良・開発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] 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フェース」、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 verification 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] 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フェース化」、情報処理学会第 52 回全国大会 (2), 4D-3, pp.167-168, 1996
- [9] Kaplan,S.J. : "Cooperative Responses 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).
- [10] Webber,B. : "Question Answering", Shapiro,S.C.(ed.), Encyclopedia of Artificial Intelligence, pp.814-822, New York: Wiley(1987)
- [11] 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) pp.353-354, 1996.
- [12] 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