

マルチモーダルなエージェント型ユーザ インタフェースの評価と対話制御の検討

安藤 ハル 畑岡 信夫

(株)日立製作所 中央研究所

〒185 東京都国分寺市東恋ヶ窪1-280

e-mail : haru@crl.hitachi.co.jp

新しいヒューマンインタフェースの試みとして、音声入出力技術を利用したマルチモーダルな入出力を持つエージェント型インタフェースの構築を行なっている。ユーザが親しみやすく、かつ対話性と操作性の良いインタフェースとするために、「音声入力」と「ポインティングジェスチャ」を入力手段とし、さらに「音声出力」と「3次元CG」にて表現されたエージェントを用いている。本報告では、まず、試作したシステムを用いてエージェント型インタフェースの有効性を確認したので報告する。さらに、エージェントの対話制御の形態が、ユーザのシステム操作方法の習得にどの様に影響するのかを評価した結果、本報告で提案する主導権移行型の対話制御が有効であることを確認した。

Study on Multimodal and Agent-typed User Interface Using Speech Dialogue and Pointing Gestures

Haru Ando and Nobuo Hataoka

Central Research Laboratory, Hitachi Ltd.

1-280 Higashi-Koigakubo, Kokubunji-Shi, Tokyo 185, Japan

e-mail : haru@crl.hitachi.co.jp

A multimodal interface has been regarded as one of the most natural and easy-to-handle man-machine interfaces for unexperienced computer users. We have implemented an "Agent-typed Interior Design System" as a prototype for evaluating input and output modalities. An agent is realized by 3 dimensional CG and speech guidance in this interface. In this paper, the details of system implementation and evaluation results, which clarified effects of the proposed agent-typed interface and effects of a dialogue control mechanism in which the dialogue initiative can be transferred from systems to users, are described.

1. はじめに

ユーザが自然にかつ容易に利用できるインタフェースとして、複数の入出力手段を用い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いる[1][2][3]。マルチモーダルインタフェースを用いることによって、ユーザは、音声は当然のことながら、ジェスチャや表情、視線などを用いて計算機に指示が行なえるようになる。また、ユーザと計算機の対話性を向上させることを目的として、エージェント型インタフェースの研究も盛んになっている[4][5][6]。エージェント型インタフェースを用いて、システム側から音声やジェスチャを用いた適切な応答を得ることによって、計算機との自由度の高い対話が可能となる。本研究においては、まず、マルチモーダルなエージェント型インタフェースとして「インテリアデザイン支援システム」を試作し、試作システムの操作性の評価を通じて、エージェント型ユーザインタフェースの有効性を確認した。さらに、エージェントの対話制御において、対話制御方法によるユーザの操作学習効率の変化を評価し、画面表示を利用してシステム主導型からユーザ主導型に移行できる「主導権移行型」の有効性を確認した。

以下、第2章で、試作したエージェント型インタフェースについての概要、及び本システムを用いたインタフェースの評価について述べる。さらに第3章において、対話制御方法である「主導権移行型」の提案と本対話制御方法の有効性の評価に関して述べる。

2. マルチモーダルなエージェント型ユーザインタフェース

2.1 インテリアデザイン支援システム

マルチモーダルな入力手段としては、音声入力、ジェスチャ認識、視線認識、表情認識などがある。出力手段としては、音声出力、3-D CGでの画像出力等がある。本研究においては、入力手段として、ユーザの能動的な入力である音声入力とポインティング指示入力を用いた。出力手段としては、エージェントの表現形態として3-D CGによるエージェント表示及び音声出力を行なう。本システムは、部屋などのインテリアを設計する場合に、「椅子」、「机」などの家具を配置したり、各々の家具の大きさ、色などをその場で変えたりするシミュレーションをワークステーション上で

行なうためのシステムである。まずシステム全体の構成を述べ、次にシステムの機能を示す。

(1) システムの構成

図1に本システムの構成を示す。ワークステーションを、音声認識用に1台、アプリケーション駆動用に1台用いている。「ポインティング」を取り込むために、ディスプレイに光方式のタッチパネルを取りつけた。音声入力処理部では、音声認識方式として、離散型HMMをベースとしたフレーム同期Viterbiビームサーチ方式を採用している[7]。表1に認識系の諸条件を示す。正規文法と辞書を用いて作成されたHMMのネットワークと入力音声の特徴ベクトルの時系列とを照合して、最も確率の高いパスを構成する単語系列が認識結果となる。認識処理は、ほぼ実時間で実行されている。この認識結果は、イーサネットを通じてアプリケーション駆動部に送られる。また、タッチパネルから入力されたポインティング情報は、RS232Cケーブルによって同じくアプリケーション駆動部に送られる。音声認識結果とポインティング情報は、アプリケーション駆動部において統合され、ユーザの意図を理解する。また、情報出力として、音声出力、インテリア編集画面、そしてエージェント表示画面を用意している。音声出力には、外部取付けスピーカを2台接続しており、出力内容は、1文を1ファイルとして、音声波形で記憶されている。インテリア編集画面は、XWindow上での描画であり、2次元である。エージェントは、3次元CGで表現されている。

(2) システムの機能

本システムでは、ユーザが画面に向かって座り、マイクで「椅子をこちらに置いて。」と言い

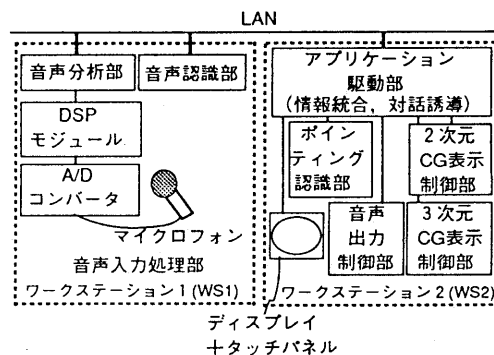


図1 インテリアデザイン支援システムの構成

ながら、指を用いてタッチパネル上で移動位置を指定する。図2にシステムの処理フローを示す。ユーザが音声とポインティングで入力を行なうと、システムはまず、入力音声の認識及び意味解析を行なう。次に、タッチパネルから入力されたポインティング情報と統合し、コマンド、対象物、条件等を抽出する。次に、入力情報判断部で、意図理解に必要な情報が満たされているかどうかを判断する。情報が満たされている場合には、コマンドを実行して対話誘導部で応答内容を選択し、応答メディア選択部で選択された出力モダリティによって情報を出力する。情報が不十分な場合には、システム側からエージェントを介して確認或いは入力要求を行ない、ユーザの不足情報入力を促進する。この機能によって、ユーザから入力される可能性のある語彙を削減することが可能になり、認識率向上にもつながる。ユーザが可能な操作は、「移動」、「複写」、「削除」、「拡大縮小」、「着色」、「交換」、「新規入力」の7種である。

2.2 操作性の評価

実際に、被験者に試作システムを使用してもらい、エージェント型インタフェースの有効性を評価した。以下に、評価目的、評価実験の概要、及び結果、考察について述べる。

2.2.1 評価目的

評価目的を以下に示す。評価においては、比較対象として、3種のシステムを用いた。第1は、エージェント機能を持たないシステムであり、出力としては、インテリアデザインの編集画面のみである。第2と第3は、エージェント型インタフェースを搭載したシステムであり、音声出力の

みのシステムと、音声出力と3次元による女性エージェントの表示を持つシステムである。

2.2.2 評価実験の概要

被験者：男性5名、女性5名

入力方法：音声で「これを削除して。」などと連続発声しながらポインティングで移動位置等を指示する。

評価実験の手順：

- 1) ディスプレイに初期レイアウトを提示する。
- 2) 紙に描かれた目標レイアウトを示し、被験者に、初期レイアウトを目標レイアウトの様に編集するように指示する。
- 3) 編集時に、上記の3種のシステムによって同様の編集過程をたどり、レイアウトを実行させる。
- 4) 各システムに対して操作性を順位づけさせる。
- 5) 評価結果の理由を記述させる。

なお、コマンド数は5個で、「移動」、「複写」、「削除」、「拡大縮小」、「着色」とした。また、実験に入る前に音声認識の練習を5回ずつ行なった。このとき被験者には、発声するときになるべく切らずに、自然に発声するように指示した。さらに、3種のシステムに対して同様の編集過程をたどらせるために、編集過程を覚えさせ、本番に臨んだ。評価項目として、(1)システムの親しみやすさ、(2)システムの使いやすさ、(3)作業効率の良さの3つの観点から評価させ、評価理由を書いてもらった。順位付けについては、評価で差がない場合には、例えば、1位として2種のシステム、3位として1システムといったように同順位をつけることを許した。

2.2.3 評価結果及び考察

評価結果を図3に示す。図は各項目についての

表1 音声認識の条件

音声分析	サンプリング	12 kHz, 16 bit
	特徴パラメタ	14次LPC ケプストラム
		LPC ケプストラムの回帰係数
		パワーとパワーの回帰係数
	分析窓長	20 ms
	分析周期	10 ms
	ベクトル量子化	各パラメタ256レベル
HMM	モデルタイプ	離散出力確率分布型HMM
	単位	CV音節 (クラスタリングで最適化)
	モデル構造	5状態 (5ループ)
サーチ	フレーム同期	Viterbi ビームサーチ

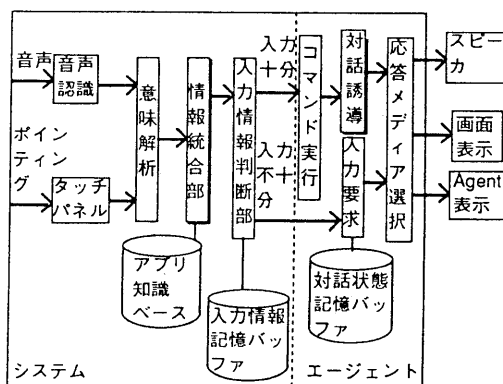


図2 エージェント型インタフェースの処理フロー

比較評価結果である。縦軸が人数、横軸が3種の評価システムを表し、各システムに対する操作性が評価されている。例えば、黒塗り部分は、各項目に対して、ユーザが最も良いと判断したインタフェースである。以下に項目ごとの結果及び考察について示す。

(1) システムの親しみやすさについて

図3から本研究で取り上げたエージェント型インタフェースが親しみやすく、特に編集画面、音声出力及びエージェント表示によるインタフェースが最も親しみやすいということがわかる。エージェント表示、対話を有効とするユーザの評価理由は以下の通りであった。

- 1) 音声で入力するため、音声で応答が返ってくることが自然に感じる。
- 2) 音声出力での状況説明があると安心できる。
- 3) 音声があると入力のタイミングがわかりやすく、対話している気持ちになる。
- 4) 音声出力だけでは、どこから聞こえてくるのが不安であり、エージェントの表示が必要。

(2) システムの使いやすさについて

図3から、音声出力、或いはエージェント表示と音声出力を用いた場合の方が、ユーザにとって使いやすいことがわかる。音声出力がシステムの使いやすさにとって重要であるとするユーザの評価理由は以下に示す。

- 1) 音声出力は、次に何をすれば良いかを教えてくれるので、初心者にとって必要であると感じた。
- 2) 音声応答やエージェントの表示によってシステムの状態がわかりやすいため不安を感じない。

(3) 作業効率の良さについて

図3から、作業効率に関しては、評価にばらつ

きがあることがわかる。この評価結果の原因は、評価実験のタスクが小さかったことにありと考える。ただし、傾向としては、やはりエージェント表示がある方が有効であると言える。ユーザの評価理由を以下に示す。

- 1) 基本的には、3種変わらないが、エージェントの音声応答やジェスチャによってシステムの様子が明確になるため効率良く感じる。
- 2) 入力の際、エージェントが入力すべき内容について音声応答やジェスチャによって誘導してくれることが疲労度を軽くするように感じる。
- 3) 慣れてくれば、音声でなく文章などの表示だけでもよいのではないかな。

以上のことから、傾向としてはエージェントがあった方が有効であるが、タスクの難しさやタスクの大きさに依存して、作業効率の良さに違いが出てくるのではないかとということが示唆された。

3. エージェントによる対話制御の評価

本章では、対話機能を有するエージェントの対話制御が、どのようにユーザの操作学習効率に影響するのかを評価し、画面表示を利用してシステム主導型からユーザ主導型に移行できる主導権移行型の対話制御方法の有効性を確認した。評価では、音声認識部に人間が介入する「Wizard of Oz」方式を用いた。

3.1 システムとユーザとの対話制御

対話における主導権をどのように制御するかは、対話機能を持つエージェント型インタフェースを構築するに当たって重要な項目である。代表的な対話制御方法としては、システムから入力内容を規定するシステム主導型と、ユーザ側からあ

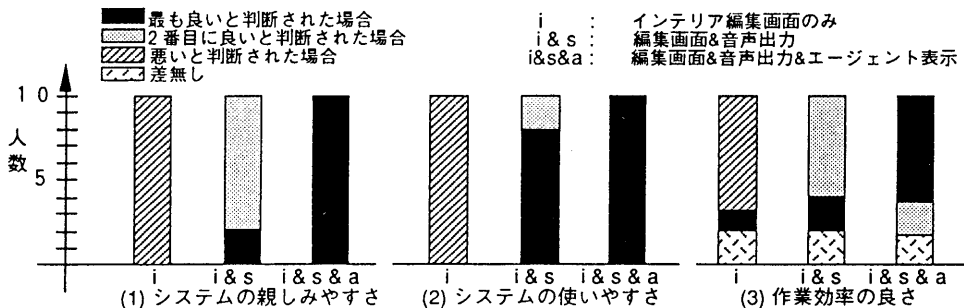


図3 エージェント型インタフェースの評価結果

る程度自由に情報を入力できるユーザ主導型の2種類がある[8]。システム主導型は、操作が不慣れたユーザにとっては有効である一方、問いかけに従ってしか入力できず、操作を学習した後の対応に問題がある。一方、ユーザ主導型は、ユーザが入力したい内容がある程度自由に入力できるが、現状の音声認識技術の制約から、全く自由に発声することはできない。従って、ユーザ主導で入力を行なう場合には、ユーザはあらかじめ入力可能な発声内容を記憶しておく必要がある。そこで、今回、システム主導型からユーザ主導型へ主導権移行が可能な対話制御方法を提案する。本制御方法では、対話がシステム主導型で始まり、各編集の実行後、画面の一部にユーザ主導型での入力方法が表示される。ユーザは、表示内容を記憶することによって、後に、同様の編集内容が出現した場合、ユーザ主導型で入力できるようになる。この結果、不慣れたユーザでも操作方法を徐々に習得でき、作業効率が上がると考えられる。

3.2 操作学習過程における対話制御評価

3.2.1 評価目的

システム主導型と、ユーザ主導型、及び主導権移行型の3種類の対話制御とユーザの操作学習効率の関係を評価する。評価実験では、音声の誤認識がインタフェース評価へ影響することを防ぐために、「Wizard of Oz」方式を用い、音声認識率を100%とした環境を設定し、評価を行なった。

3.2.2 「Wizard of Oz」方式での実験環境

実験システムの構成を図4に示す。2名の実験者を用意し、実験者1は、被験者に対して実験の説明を行ない、実験者2は、発声内容をシステムの代わりに認識し、操作を行なうWizardである。被験者は、防音室内で被験者用ワークステーション2(WS2)のディスプレイを見ながらタスクを遂行する。Wizardは、防音室外に位置し、被験者に操作を気づかれないようにした。Wizardは、設置したテレビに送られてくる被験者の行動を映した映像と音声を受けることによって被験者の入力内

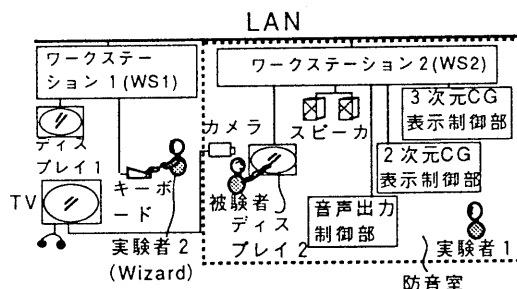


図4 「Wizard of Oz」方式の実験環境

容を理解し、実験者用ワークステーション1(WS1)のキーボードからコマンド、マウスから被験者の指示位置を入力する。その結果、WS1、WS2の画面に編集後の同一画面が表示される。

3.2.3 評価実験の内容

- ・被験者：男性5名、女性5名
- ・タスク：タスクの難易度を反映した易しい、小さなタスクと、難しい、大きなタスクを用いた。小さなタスクでは5試行、大きなタスクでは20試行の編集を行なう。
- ・対話制御方法：対話制御は、(1)システム主導型、(2)ユーザ主導型、(3)主導権移行型の3種である。対話の進行を図5に示す。(1)は、システムの問いかけに対してユーザが対象物名等を入力する。(2)は、各編集内容に対して適当な発声内容を表から選択して入力する。(3)は、最初はシステム主導型で対話が始まるが、各編集実行後に画面の一部にユーザ主導型での入力方法が表示され、入力の学習ができる。

#	対話制御	編集内容	システムの出力	ユーザの発声
(1)	システム主導型	移動-1	対象物はどれですか？ 対象物をどうしたいですか？ どこに移動しますか？	“これ” “移動したい” “ここ”
(2)	ユーザ主導型	移動-1	かしこまりました。	“これをここに移動して”
(3)	主導権移行型	移動-1 (システム主導型)	対象物はどれですか？ 対象物をどうしたいですか？ どこに移動しますか？	“これ” “移動したい” “ここ”
		移動-2 (ユーザ主導型)	“画面表示” ユーザ主導型での入力方法： “これをここに移動して”	“これをここに移動して”

図5 各対話制御の過程

・実験手順：まずタスク表に示されている内容に従って画面を編集するように指示する。次に、3種の対話制御方法を各々用いて、実際に音声とポインティングを使って画面の編集をしてもらう。最後に、3種の対話制御に対する操作性の順位づけを行なわせ、評価結果の理由を各対話制御毎に記述させる。

3.2.4 評価結果及び考察

評価結果を図6及び図7に示す。それぞれ小タスク、大タスクにおける評価結果である。縦軸が人数、横軸が対話制御の種類を表し、各対話制御に対する操作性が評価されている。評価結果及び考察は以下の通りである。

1)小タスクについて：図6から、システム主導型が最も操作性が良いことがわかった。被験者の評価理由は以下の通りである。

・システム主導型は、時間はかかるが、入力の間違いが少ないと感じた。

・ユーザ主導型は、入力できる内容を用紙チェックしてから入力しなければならないので実際の作業が中断される。

2)大タスクについて：図7から主導権移行型が優位であることが分かった。被験者の評価理由を以下に示す。

・主導権移行型は、操作を覚えれば使い易い。

・主導権移行型は、操作が不慣れな場合、システム主導で確認しつつ指示できるので操作しやすい。

但し、エージェントの入力要求に引き込まれ、結局システム主導になってしまうという意見もあり、主導権を移行するタイミングが難しいことがわかった。

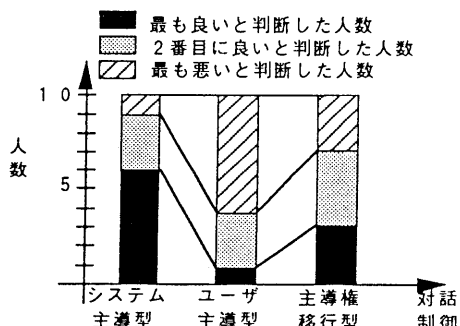


図6 小タスク (5 試行) の場合の操作性評価結果

4. まとめ

本報告では、ユーザとシステム間の円滑な対話を実現するために、まず、マルチモーダルなエージェント型インタフェースの対象システムとして「インテリアデザイン支援システム」を試作し、本システム上でインタフェースの有効性を評価した。評価の結果、エージェント型インタフェースが、システムの親しみやすさ、使いやすさ等、操作性向上の観点から有効であることを確認した。さらに、エージェントの対話制御の形態が、ユーザのシステム操作方法の習得にどのように影響するのかを評価し、提案した「主導権移行型」の対話制御方法が有効であることを確認した。

参考文献

- [1] 新田：マルチメディアとマルチモーダルUI, 情報処理学会音声言語研究会資料, 1-3, pp.31-38, (1994)
- [2] Nigay.L & Coutaz.J : A Generic Platform for Addressing the Multimodal Challenge, CHI '95, pp.98-105, (1995)
- [3] 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, 信学論D-II, Vol.J77-D-II, No.8, pp.1465-1474, (1994)
- [4] Macs. P : Learning Interface Agents, FRIEND21 '94, International Symposium, (1994)
- [5] 伊藤, 長谷川, 速水, 田中, 山本：エージェント型マルチモーダルインタフェースの試作, HIシンポジウム, pp.309-312, (1994)
- [6] 安藤, 畑岡：音声とポインティングジェスチャを入力手段としたマルチモーダルインタフェース, 95年春季信学会全大 SD-9-1, (1995)
- [7] 小高, 天野：電話回線とLANを介した音声認識応用の検討, 信学技報SP94-55, pp.15-21 (1994)
- [8] 山本, 中川：音声対話システムの構成法とユーザ発話の関係の模擬実験による評価, 情報処理, Vol.95, No.16, pp.61-68, (1994)

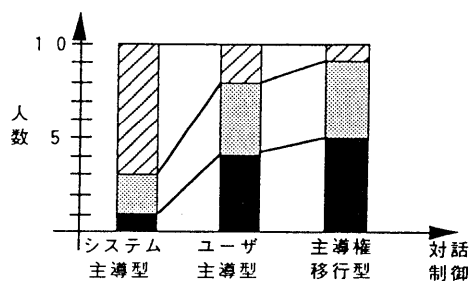


図7 大タスク (20 試行) の場合の操作性評価結果