

## 7H-6 擬人化エージェントにおける WWWブラウザを用いたマルチモーダルインタフェース

久保田 仙 高間 康史 土肥 浩 石塚 満

東京大学 工学部 電子情報工学科

kubosen@miv.t.u-tokyo.ac.jp

### 1 はじめに

現在、社会の多くの場所にコンピュータが普及し、程度の多少を問わなければ、ほとんどすべての人がコンピュータと接して生活しているが、現状では人間側がコンピュータの都合に合わせて使用しており、コンピュータを扱うには知識や経験が不可欠である。もしもこのような前段階が不要となれば、初心者も手軽にコンピュータを扱えるようになる。一方、コンピュータになれた人でも、コンピュータとのより幅広いインタラクションが可能になれば、コンピュータをより有効に活用でき、多様な活動を行えると考えられる。そこで、当研究室では自然感の高い顔画像を有し、音声対話によってユーザとのコミュニケーションを行なう擬人化エージェント（VSA）というシステムが開発されている [2]。さらに、音声による応答と、データを表、絵、文を組み合わせるなどして、複数の知覚（マルチモーダル）を用いてユーザへ情報を提供する機能が追加され、デパートの案内を目的としたシステムの開発が行なわれた [1]。また、このシステムには、ユーザとの対話を通じてより協調的な対話の実現を目指すための学習機能が付与されている [4]。これに加えて、本研究では、現在広く普及しつつあるWWWブラウザを利用してマルチモーダル応答を実現する。Netscapeなどの既存のWWWブラウザを利用することにより、マルチモーダルインタフェースの開発が容易になるだけでなく、従来のマウスのみを用いたブラウジングよりも柔軟なインタラクションが可能となることが期待される。

### 2 マルチモーダルな応答とは

モダリティとは、視覚、聴覚、触覚などの感覚を用いて外界からの情報を知覚するやり方であり、人間同

士の会話には言葉以外に身振りなどのさまざまなモダリティが自然に使われている。人間とコンピュータ間のインタラクションもまた情報の送り手と受け手が複数のモダリティを用いて相手の意図を解釈しながら対話する、マルチモーダルインタラクションがこれからの望ましい姿であるといえる。

ユーザインタフェースにモダリティの相補性を利用する目的は、次のようなものがあげられる。

- 複数のモダリティを組み合わせることにより、ユーザインタフェースとしての頑健さをたかめる。
- 複数のモダリティを組み合わせることによって意味の確定、強化をする。
- ユーザの認知的負担や心理的抵抗感を軽減する。
- 状況に応じたモダリティの組合せによってわかりやすい応答を作成できる。

画像情報は大量の情報を一度に提示できる利点があるが、データベースの検索結果全てを無制限に羅列するのでは協調的な応答とはいえない。そこで本研究ではデータベースの検索結果を元に表示形式を動的に決定することに焦点をあてる。

### 3 システム概要

本研究は、文献 [1] によって報告されているマルチモーダル応答システムをベースとし、WWW上で、多くの情報から適当なものを選びだし、見やすく表示案内するようなエージェントの作成を目標とする。その概要は以下の通りである。

まず、ユーザの発話からキーワードスポッティングを用いて意味同定に必要なキーワードが抽出する。次に、得られたキーワード列から、その発話が、どのような意図を持って質問されたかを決定し、意味表現へと変換する。これまでを状態認識部とする。

状態認識部で決定された意味表現は、次に応答決定部に送られ、そこで質問に応じた必要なデータを検索用のデータファイルから選出し、そのデータは応答作成部へと送られる。

応答作成部では、応答決定部から送られてきたデータの量に応じてそれに見合った形式で応答を作成する。画面の出力はHTMLファイルを作ってブラウザに、音声出力は出力装置にそれぞれ送る。

#### 4 システムの応答作成例

現在、システムのタスクとして、デパートの食品フロアの案内をとりあげている。このようなシステムを利用するユーザはあらかじめ明確にゴールを決定しているとは限らないので、質問を繰り返して、フレームを埋め込んでいくような対話形式をとった場合には、十分な検索キーが得られないことも多いであろう。この場合、ユーザに尋ねられた質問に答えるような会話をしながらユーザとともに必要な情報を探しだしていくのが望ましいが、本研究では、マルチモーダルな応答の性質を利用して、これを実現する。

具体的には質問に応じてデータベースを検索した結果、得られたデータのエントリ数によってブラウザに表示するデータの種類（テキストや画像など）、表示形式（リスト、表など）を変化させる。また、強調すべき情報を音声によって出力することにより、ユーザの注意を促す。グラフィックデータの配置や、表の作成などには、WWWブラウザの能力を有効に利用する。

以下に応答作成の例を挙げる。

- 提示すべきエントリ数がかなり大量にある場合：画面には表を作成して表示する。音声では”御覧のようになります。”などととどめる。
- 提示すべきエントリ数が3-5の場合：画面には表といくつかの強調すべき絵を表示し、音声は、すべてを読み上げる。(Fig2.1)
- 提示すべきエントリ数が1-2の場合：画面には、それぞれの絵や、説明、さらに関連する情報を補い、音声でその説明をする。(Fig2.2)

#### 5 まとめ

本研究では、案内などを行なう擬人化エージェントの開発にあたり、WWWブラウザを利用してマルチモーダルな応答を作成し、また、提示する情報量によってその応答内容を変化させることによって、より協調的な応答が実現出来るということについて述べた。今後はこのシステムにさらに学習機能をつけるなどして、さらに充実した応答の生成を目指す。

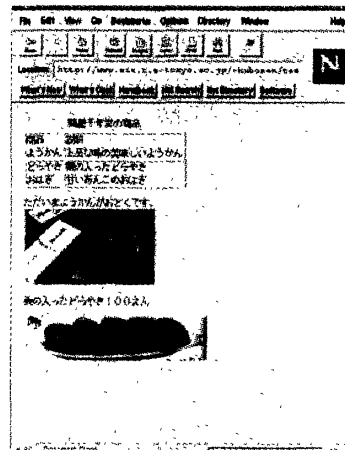


Fig2.1 回答画面例

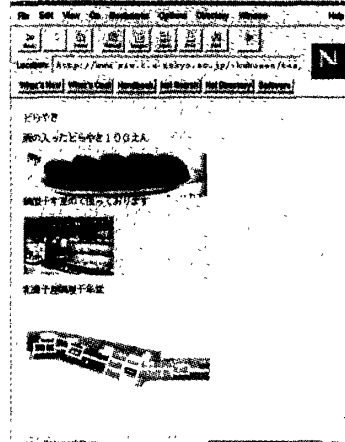


Fig2.2 回答画面例

Fig2. 回答画面例

#### 参考文献

- [1] 原田 晋；擬人化エージェントにおけるマルチモーダルな案内応答の作成；1995年度卒業論文
- [2] 土肥, 石塚；WWW / Mosaic と結合した自然感の高い擬人化エージェントインタフェース；電子情報通信学会論文誌 D-II Vol.J79-D-II No.4 pp.585-591
- [3] 長尾 確；マルチモーダルインタフェースとエージェント；人工知能学会誌 Vol.11 No.1 p.32-40 1996
- [4] 高間, 土肥, 石塚；擬人化エージェントにおける音声対話を通じての協調的応答戦略の自動学習；人工知能学会誌 Vol,12 No,2 1997 (予定)