7H－7

擬人化エージェントを用いたインターネット上のマルチモーダル情報ガイドシステム

森 真史　土肥 浩　石塚 満
東京大学　工学部　電子情報工学科
mori@miv.t.u-tokyo.ac.jp

1　はじめに

従来より我々は新時代のヒューマンインタフェースの形態として擬人化エージェントシステムを用いた対話システムを研究してきた。これはコンピュータ利用者からの音声入力に対してコンピュータ画面上の顔画像と音声が実時間で応答するというものである。

使いやすいヒューマンインタフェースを実現するためには様々な手法が考えられるが、より人間同士のインタフェースに近づけるという観点から見ればコンピュータ上に仮想的な人間であるところの擬人化エージェントを生成する対話システムというアプローチは広い層の利用者にとって違和感がなく親しみやすいものである。

さらにこのヒューマンインタフェースの形態は慣れによる操作の熟練が生じる場面、すなわち同一人が日常的に利用する環境よりむしろ、初めての利用者が多数である情報ガイドシステムにおいてより効果的であると考えられる。

他方、我々の社会においてはここ数年でWWWを中心としてインターネットが一般に広く利用されるに至り、コンピュータの果たす役割は作業の効率化という無機的なものから情報発信という人間的なものへと激しく変化しつつある。

そこで、我々はこの擬人化エージェントシステムを用いてインターネット上にマルチモーダルな情報ガイドシステムを構築することを考案した。

以下ではこのシステムの概要について述べる。

2　システム構成

システムはサーバ/クライアント構成で、既存のWWWシステムへの拡張の形態をとる(図1)。

サーバは画像生成、画像制御、音声生成、音声認識、対話管理の各ユニットからなり、クライアントからの要求に応じてCGIスクリプトとして起動される。

クライアントは画像保存、画像再生、音声再生、音声録音の各ユニットと、スピーカ、マイクなどのハードウェアからなり、WWWブラウザから起動される。ただし、クライアント起動後はWWWブラウザとは独立して動作する。

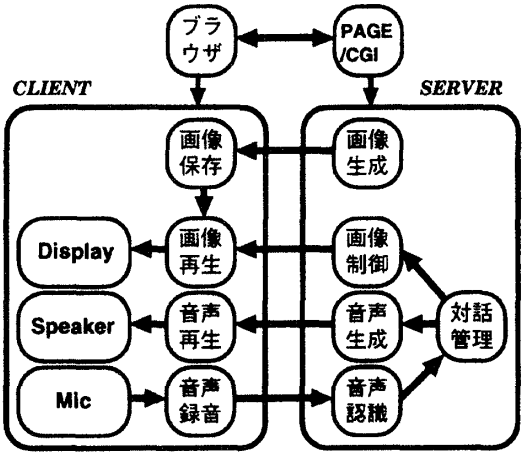

図 1: システム構成

サーバ/クライアント間の通信はWWWと同じIPポートを介して行なわれる。

3　特徴

本システムには以下のような特徴がある。

- 発信者側が意図した通りの外観、声音を持つ擬人化エージェントの実現

  音声生成部および、顔画像生成部をサーバ側に保有することで発信者側が意図した通りの外観および声音を持つ擬人化エージェントを実現することが可能となる。

  これは特に宣伝広告など感性に訴える情報を発信する際には重要となる点である。

- 既存のWWWシステム、HTTPの枠組の利用

  本システムを既存のWWWサーバやWWWブラウザに対する拡張ユニットとして実現することにより、現在すでに広く普及しているWWWシステムの資源をそのまま利用することが可能となる。

- 複数のクライアントとの同時接続の実現

Masafumi MORI, Hiroshi DOHI, Mitsuru ISHIZUKA
Dept. of Information and Communication Engineering,
Faculty of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113, JAPAN

サーバ機能をソフトウェアのみで実現することにより、単一のサーバにおいて同時に複数のクライアントからの要求を処理することが可能となる。

- 簡易な機器構成によるクライアントの実現

  音声認識、音声生成、画像生成などの特殊な機能あるいは大きな計算量が要求される部分をサーバに用意することで、クライアントを一般的で安価なシステムによって実現することができ、より多くの利用者を見込むことが可能となる。

4 実装

システムを実装するにあたってはデータ圧縮アルゴリズムやデータ転送方式など各部分に対して様々な実装方式が考えられるが、現在の実装においては以下の方式を試みている。

4.1 処理手順

本システムでは以下の手順で処理が行なわれる (図2)。

1. 通常の WWW システムから本システムが起動される
2. 接続開始時にサーバからクライアントへ擬人化エージェントの要素画像などを転送する
3. サーバからクライアントへ質問の音声データと顔画像制御用の信号とを転送する
4. クライアントは音声を再生すると同時に顔画像を動作させる
5. ユーザはシステムからの質問に対する答えを発声する
6. クライアントは録音した音声をサーバへ転送する
7. サーバは音声認識を行ない適切な応答を決定する
8. 上記 3〜 7を反復する

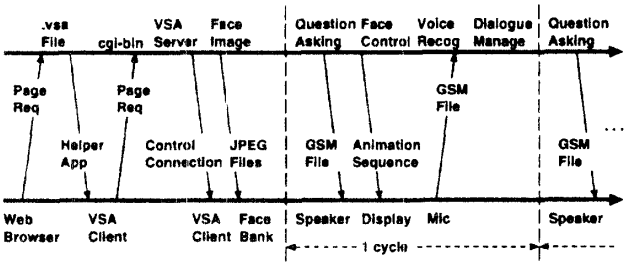

図 2: 処理手順

4.2 画像

擬人化エージェントは現実感のある顔画像をリアルタイムで表示する必要があるが、サーバ / クライアントのいずれにその機能を持たせるかによって通信容量と計算量のトレードオフが存在する。そこで、あらかじめサーバで要素画像を作成、圧縮しておき、接続開始時にクライアントにすべて転送した上で、サーバの指令に従ってクライアントが画像を伸長、表示する方式を採用する。

ここでは画像圧縮アルゴリズムとして JPEG を用いており、240×320pixel、24bit/pixel の画像で概ね 15kbyte 程度である。

4.3 音声

音声情報に対する圧縮アルゴリズムとして GSM 06.10 を用いており、データ量は 13kbit/s となる。また、音声データの転送には IP パケットを採用したが、これは用いる音声データが小さく、遅延がそれほど目立たないためである。

音声認識部は不特定話者による単語認識が可能である。また、システムからの質問内容によってユーザの返答を予測し、その結果に基づいて音声認識辞書を動的に再構築し、認識範囲を絞り込むことで認識効率を向上させることができる。

5 おわりに

インターネット上で擬人化エージェントと対話する情報ガイドシステムは、ホームページの訪問者に対する電子秘書や、大学キャンパスやデパート売場での案内人など幅広く利用できる技術であると言える。

なお、本システムの実装においては音声認識システムとしてオムロン株式会社の FI 音声認識技術をご厚意により使用させていただいている。

参考文献

[1] M. Mori, H. Dohi and M. Ishizuka: "A Multi-Purpose Dialogue Management System employing Visual Anthropomorphous Agent", Proc. ROMAN'95, pp.187–192, Tokyo (1995.7)

[2] 森 真史, 土肥 浩, 石塚 満: "擬人化エージェントを用いたインターネット上のマルチモーダル情報ガイドシステム実現機構", 情処 52 回全大, (6)pp.201–202, (1996.3)