

3 Q - 5

WWWを情報源として音声応答する
擬人化エージェントインタフェース土肥 浩 石塚 満
東京大学工学部電子情報工学科

1 はじめに

インターネット上のワールドワイドウェブ (WWW) では、多くのユーザにより多種多様の情報が提供され、またその内容は常時更新されている。この WWW を分散型大規模情報データベースとみなして、天気予報・降水確率やプロ野球/Jリーグ/大相撲の結果など、ユーザの質問に対して音声応答する擬人化エージェントインタフェースを実現した。エージェントはユーザの指示とは独立に、複数のサブエージェントを用いて自発的に WWW サーバにアクセスし、必要なデータを収集する。これにより、擬人化エージェントのローカルデータベースの内容は、人手を介さず、自動的に更新される。

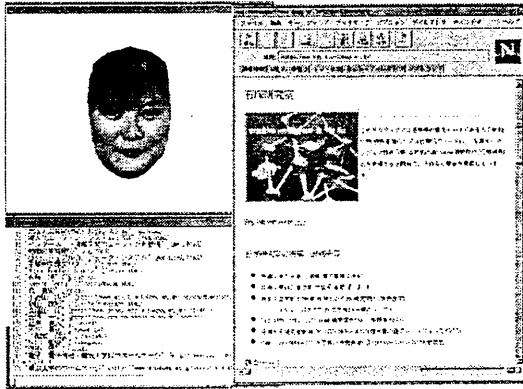
2 WWW ブラウザと結合した
VSA インタフェース

図 1: VSA interface connected with Netscape

我々は、自然な人間の顔を持ち、インタラクティブにユーザと音声対話できる知的擬人化エージェントであるビジュアルソフトウェアエージェント (VSA: Visual

A Visual Software Agent Interface with Speech Dialog Function based on WWW

Hiroshi DOHI and Mitsuru ISHIZUKA
University of Tokyo, Faculty of Engineering, 7-3-1, Hongo,
Bunkyo-ku, Tokyo 113, JAPAN
(E-mail: dohi@miv.t.u-tokyo.ac.jp)

Software Agent) の研究を進めている。

これまでに、この VSA をインターネット上の WWW ブラウザ NCSA Mosaic と結合した。[1, 2] 現在では、WWW ブラウザとして Netscape Navigator を用いるバージョンも実現している。Netscape と接続した VSA インタフェースの実行例を図 1 に示す。左上画面が擬人化エージェント VSA、左下画面は各 web ページから自動的に抽出したアンカーリストである。

システム構成を図 2 に示す。VSA サーバは、規則音声合成および音声に連動した自然な顔の CG 合成を行う。インタフェースマネージャは、音声認識した結果に基づいて VSA サーバと WWW ブラウザを制御する。また音声応答などのためのローカルデータベースを管理する。音声認識には、市販の不特定話者連続音声認識装置を使用した。

ユーザは従来のマウス操作に加えて、擬人化エージェント VSA との簡単な音声対話 (例えばアンカーをクリックする代りに、その文字列を読み上げる) によって Netscape や Mosaic を制御し、WWW サーバから必要な情報を引き出すことができる。マウスによる操作と音声による操作は、完全に同じレベルで扱われる。したがって、ユーザは状況に応じてより適当なインタフェースを選択することができる。両者を自由に混在させても構わない。また音声入出力インタフェースを使用しない場合でも、Netscape や Mosaic は普通に実行される。プロトタイプシステムは、大ホールでの講演のようにマウスの使用が困難な環境においても VSA インタフェースが有効であることを示した。

3 サブエージェント

WWW ブラウザでは、文字列やイメージ等のアンカーをクリックすることにより、それにリンクされた情報を得ることができるように設計されている。ところがこれは、アンカーをクリックすればユーザが即座に必要な情報を入手できることを必ずしも保証するものではない。アンカーのクリックは、WWW ブラウザに対して WWW サーバからリンクされたデータをダウンロードするように指示する。しかしネットワークの状態によっては、そのデータがユーザに提示されるまでには長い時間を必要とするかもしれない。

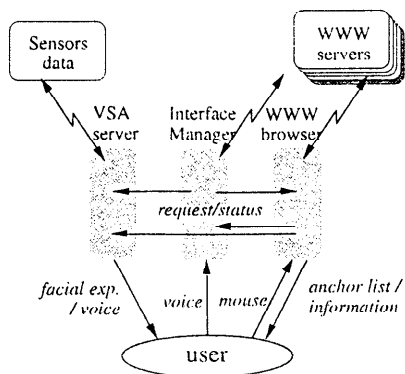


図 2: The design of VSA interface

ユーザが本当に必要としている情報は web ページの中のごく限られた一部分で、言葉にすれば一言で伝えられるという場合も多い。例えば「明日の天気を知りたい」という場合、それは一般に、自分のいる場所の天気予報や降水確率を知りたがっていると考えるのが自然である。細いネットワークで大量の画像データを転送し、長い時間をかけて全国各地の天気を美しく表示することがいつも有効であるとは限らない。

音声対話システムでは、ユーザの質問に対して遅滞なく反応（応答）できることが重要である。ユーザから質問を受ける毎に WWW サーバをアクセスしていたのでは、円滑に対話を進めることができない。そこで音声応答に必要なデータの一部は、あらかじめ特定の WWW サーバにアクセスして、その web ページから切り出しておく。

新聞紙面のレイアウトは、写真の配置や見出しの付け方など、毎日異なっている。一方、web ページの場合、それぞれのページ毎に比較的決まったレイアウトを採用している場合が多い。そのため、人手を介さずにデータを切り出すのに都合がよい。HTML で記述された web ページは、様々なタグ情報を含んでいるが、基本的にはテキスト文書である。したがって、もし web ページのレイアウトが決まっているならば、普通の文字列検索により、天気予報、前日のスポーツの結果、あるキーワードに関連するニュースなどを自動的に拾い出すことは難しくない。

この作業をバックグラウンドで専門に行うプロセスを、サブエージェントと呼ぶ。サブエージェントは、ある WWW ページから特定のキーワードを含むパラグラフ、あるいは特定のデータを自動的に切り出すための一種の情報フィルタリングプログラムである。現在

は、それぞれの web ページのレイアウトに依存するプログラムになっている。切り出されたデータは、ローカルデータベースに蓄積される。このデータは音声応答だけでなく、例えば図 3 のようなインフォメーションディスプレイにも使用される。上段には各ユーザの新着メール情報が、中段には天気予報と、棒グラフで示された降水確率が表示されている。

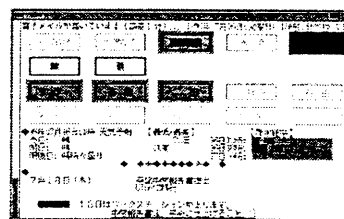


図 3: An example of information display

プロトタイプシステムでは、複数のサブエージェントが稼働している。例えば天気予報データは、1日3回、サーバ側のデータ更新に合わせて自動的にアクセスされる。またプロ野球などの試合結果は、1日1回、早朝にアクセスされる。このシステムはすでに6ヶ月以上にわたり、いくつかの WWW サーバをアクセスしてローカルデータベースの更新を自動的に行っているが、大変良好な結果を得ている。

4 まとめ

本論文では、WWW を利用して音声応答用のローカルデータベースを自動的に更新する擬人化エージェントインタフェースを実現した。天気予報・降水確率や前日のスポーツの結果など、ユーザの質問に対して音声で応答する。本研究の一部は、文部省科学研究費補助金奨励研究 (A) (課題番号 08780246) による。

参考文献

- [1] 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」, 信学論, J79-D-II, 4, pp.585-591, 1996
- [2] H. Dohi and M. Ishizuka: "A Visual Software Agent connected with the WWW/Mosaic", *Int'l Symp. Multimedia Systems (MMJ '96)*, pp.392-397, 1996
- [3] H. Dohi and M. Ishizuka: "A Visual Software Agent: An Internet-Based Interface Agent with Rocking Realistic Face and Speech Dialog Function", *AAAI-96 Workshop on Internet-Based Information Systems*, 1996