

3W-9

## ページエージェント：ホームページに棲む

擬人化エージェント  
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

Web ページを移動する度に、そのページに関係するエージェントが画面上に現われて、ガイドしてくれる擬人化エージェントインタフェース“ページエージェント”を実現した。例えば「〇〇からのメッセージ」のようなページをオープンすれば、〇〇さんの顔写真から合成したエージェントが画面に現れ、そのページの内容を音声で伝えることができる。エージェントは瞬きをしたり、喋ったりする。ホームページの作者が、そのページを紹介したりメッセージを伝えたりするエージェントを WWW サーバ側で「任命」できる。

## 2 Netscape と結合した VSA

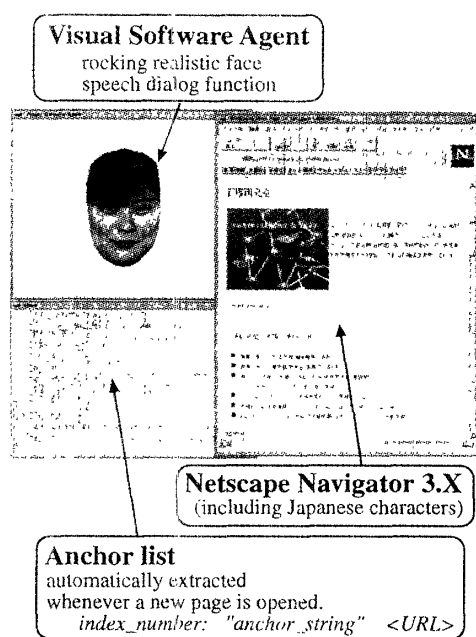


図 1: VSA interface connected with Netscape

我々はすでに Netscape や Mosaic と結合した擬人化エージェントインタフェース VSA (Visual Software Agent) を実現している [1][2]。ユーザは、アンカー文

Page Agent: A Visual Software Agent lives in a Homepage  
Hiroshi DOHI and Mitsuru ISHIZUKA  
University of Tokyo, School of Engineering, 7-3-1, Hongo,  
Bunkyo-ku, Tokyo 113, JAPAN  
(E-mail: dohi@miv.t.u-tokyo.ac.jp)

字列を発話するなど、VSA との簡単な音声対話によりインターネット情報空間を散策できる。また電子メールの到着や WWW データに基づく最新の天気予報を音声で尋ねることもできる。エージェントは、WWW ブラウザのデータ表示を妨げないように、別のウィンドウに表示される。ブラウザを見ている間は、エージェントは次の音声指示に備えて待機していた。

擬人化エージェントインタフェースには、アニメーションによる顔を用いる場合とテキストチャマッピングによる顔を用いる場合がある。テキストチャマッピングによる顔は自然感が高く、容易に人物を替えられるという特徴がある。ただし従来のシステムでは、エージェントの顔を動的に切り替えることはできなかった。

VSA では、顔の 3 次元ワイヤフレームモデルに実際の顔写真をテキストチャマッピングすることにより、エージェント顔画像を生成する。正面からの顔写真が、1 枚だけあればよい。そこでこの顔写真を WWW サーバ側に置き、任意の Web ページと関連づけることにより、ページ毎にエージェントの顔を動的に切り替えることができるページエージェントを実現した。

従来のインタフェースエージェントは、同じ人がいつも一緒に行動しながら、質問に答えたり世話をしてくれるツアーガイドのような存在であった。これに対してページエージェントは、それぞれ自分の担当するページの内容について詳しい、そこを訪れた客の応対をする店員のような存在である。

## 3 ページエージェント

## 3.1 VSA タグ

ページエージェントを指定するには http ファイルの中に、次に述べる VSA 属性ファイルの URL を 1 行記述するだけである。これは「VSA タグ」として実現されている。WWW サーバ側で、任意のページに自分の好きな擬人化インタフェースエージェントを割り当てることができる。

もしクライアントが VSA システムであれば、VSA タグを含んだページをオープンすると同時に、自動的にページエージェントが現れる（切り替わる）。そのページが表示された後、特定のアンカーをクリックすればエージェントが現れるわけではない。

VSA タグを含んだページを直接、Netscape でアクセスしても全く問題はない。その場合、ページエージェント設定のない普通の Web ページと同じに扱われる。Netscape は、VSA タグを未知のタグとして無視するからである。

### 3.2 VSA 属性ファイル

VSA 属性ファイルは拡張子“.vsa”を持つファイルで、以下の項目を含んでいる。

- 顔画像ファイル (JPEG 圧縮) の URL
- 言語 (日本語/英語)
- 音声属性 (男声/女声, ピッチ, 抑揚等)
- メッセージテキスト 1 (初回)
- メッセージテキスト 2 (2 回目以降)
- そのページに特有の (アンカーリスト以外の) 音声キーワード - URL 対応表
- 顔形状データ

### 3.3 実装

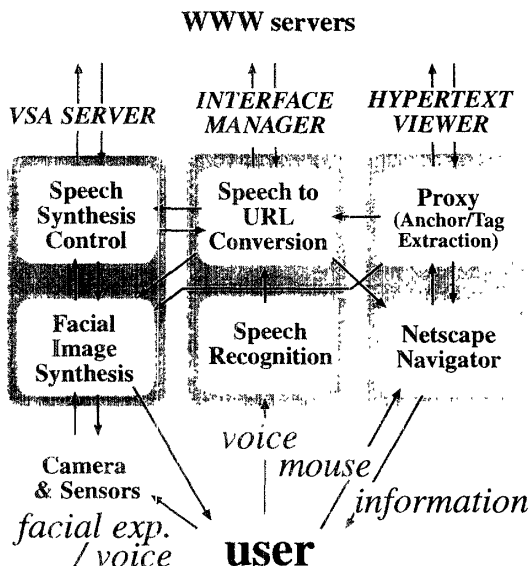


図 2: VSA implementation

ページエージェントは、VSA インタフェースシステムおよび従来の WWW の枠組みの上に実装されている。http サーバや Web ブラウザには手を加えていないので、稼働中のシステムがそのまま使える。

VSA インタフェースシステムは、Netscape を含めて六つ、あるいはそれ以上のプロセスで構成される。プロセス間は、TCP/IP または共有メモリで接続されている。各プロセスは、複数のワークステーション上で、並行あるいは並列に動作する。

VSA Proxy プロセスは、アクセスした Web ページからアンカーリスト (アンカー文字列と URL の対応表) を自動抽出する。VSA タグの処理も、ここで行われる。

1. VSA Proxy プロセスは VSA タグを検出すると、VSA 属性ファイルの URL を VSA サーバ (顔画像生成プロセス) に転送する。
2. VSA サーバは、そのデータがすでにキャッシュされているかどうかをチェックする。
3. もし初めてのエージェントであれば、Netscape のページアクセスと並列に WWW サーバから VSA 属性ファイルをダウンロードする。さらに、VSA 属性ファイルに記述された顔画像ファイルを WWW サーバからダウンロードする。
4. 伸張した顔画像ファイルと顔形状データから、擬人化エージェント顔画像を生成する。
5. 規則音声合成装置に、エージェントの音声属性を設定する。
6. 音声キーワード - URL 対応表を、インタフェースマネージャ (音声キーワード - URL 変換プロセス) に転送する。
7. 規則音声合成装置と連動して、エージェントがテキスト 1 (初回) あるいはテキスト 2 (2 回目以降) をユーザに話す。

プロトタイプでは、 $640 \times 480 \times 24$  ビットの顔画像を使用した。JPEG 圧縮後の顔画像ファイルと VSA 属性ファイルのサイズは、合わせて約 50~60KB である。顔画像は顔の部分だけがあればよいので、実際にはさらに小さくできる。

また、二つのファイルの内容はキャッシュされる。このうち顔画像ファイルのキャッシュは 2 レベルあり、LRU アルゴリズムにより最近使われた数枚は伸張した状態で、それ以外は JPEG 圧縮した状態でキャッシュされる。例えば“Back”で前のページに戻る場合、キャッシュしたデータが使われるのでネットワークオーバーヘッドはない。また JPEG 伸張も必要ないので、一瞬でエージェントの顔が切り替わる。

## 4 まとめ

エージェントとの音声対話によりインターネット情報空間を散策できる VSA インタフェースシステムをベースとして、Web ページを移動する度にそのページに関するエージェントが画面上に現われて、ガイドしてくれるページエージェントを実現した。

## 参考文献

- [1] H. Dohi and M. Ishizuka: “A Visual Software Agent: An Internet-Based Interface Agent with Rocking Realistic Face and Speech Dialog Function”, *AAAI tech. report 'Internet-based Information Systems', WS-96-06*, 1996
- [2] H. Dohi and M. Ishizuka: “Visual Software Agent Interface with Realistic Face and Voice-controlled Netscape”, *Proc. ICCIMA-97*, 1997