

WWWページごとに顔が替わる 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

WWW ページを移動するたびに、そのページに係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を実現した。WWW 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の顔や個人属性をWWWサーバ側で指定できる。ユーザは情報提供者の「顔」を見ることができる。例えば「社長あいさつ」のページを開けば、それに連動してWWWブラウザの隣りにいる擬人化エージェントが社長の顔に自動的に替わり、そのページの内容をしゃべったり簡単な音声による質問に答えたりすることができるようになる。

2 Netscape と結合した VSA

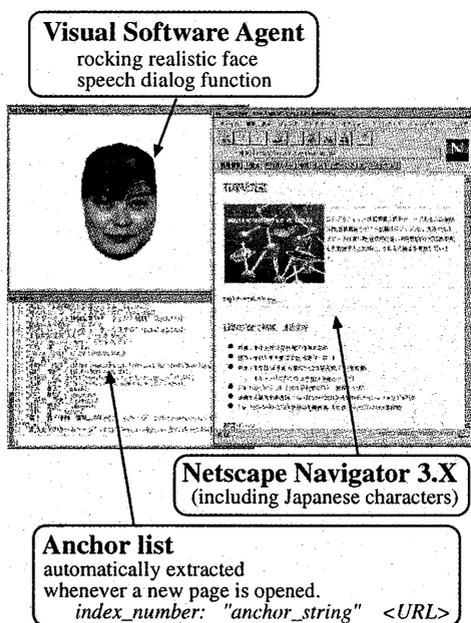


図 1: VSA interface connected with Netscape

ビジュアルソフトウェアエージェント VSA (Visual Software Agent) は、自然な人間の顔を持ち、インタラクティブにユーザと対話できる知的擬人化インタフェースエージェントである。人間の日常的な対面型 (face-to-face) コミュニケーションスタイルに基づいて、顔写真をもとに CG 合成されたエージェントとテレビ電

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)

話でおしゃべりするような感覚で、操作方法を覚えたり練習したりすることなしに誰もが高度な計算機の能力をフルに活用できるインタフェースの実現を目指している。

VSA はすでに Netscape や Mosaic と結合されており、これらの WWW ブラウザ画面を通してさまざまな情報をユーザに提供する [1][2]。これにより VSA で提供するコンテンツデータを、HTML 言語をはじめとして WWW で標準的に使用されているフォーマットで記述できるだけでなく、これまでに WWW 上で蓄積されてきた膨大なマルチメディアデータを WWW と同じように VSA から利用することができるので応用範囲が広がる。

ユーザは、通常のマウスやキーボードによる操作に加えて、アンカー文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 から自動的に取得した最新の天気予報などを、音声で聞くこともできる。マウス/キーボードによる操作と音声による操作は同じ優先度をもつため、ユーザは状況や環境にあわせて、いつでも最適なインタフェースを選択できる。

3 ページエージェント機能

3.1 ページエージェントの生成

擬人化エージェントの顔は、音声入力時にコンピュータという箱に向かって話しかけるという心理的抵抗感を和らげる。これには、TOSBURG-II (東芝) やユニバーサル電子図書館 (NEC) にみられるアニメーションによる顔を用いる場合と、VSA のようにテキストチャマッピングによる顔を用いる場合がある。それぞれ一長一短があり、どちらが優れているとは一概にはいえない。後者の特徴の一つとして、もとなる写真を交換するだけで容易に顔の印象を替えられることがある。

VSA では、正面から撮影した 1 枚の顔写真を 3 次元頭部ワイヤフレームモデルに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。そこでこの顔写真、個人属性、および音声対話用データを WWW サーバ側に置き、WWW 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が自動的に切り替わる擬人化エージェントを実現した。これをページエージェント機能と呼ぶ。

従来のインタフェースエージェントは、ツアーガイドのように同じ人がいつも一緒に行動しながら、質問に答えたり助言をしてくれる存在であった。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、そこを訪れた客の対応をする店員のような存在である。

3.2 VSA タグ

WWW サーバ側で、自分の好きな顔の擬人化エージェントを任意の WWW ページと関連づけるために「VSA タグ」を導入した。VSA タグは、次に述べる VSA 属性ファイルの URL を含んでいる。これを HTML ファイル中に、1 行記述するだけでよい。

もしクライアントが VSA システムであれば、Netscape 上のマウス操作あるいはエージェントとの簡単な音声対話により、VSA タグを含んだ WWW ページを開くと Netscape 画面にその内容が表示され、それに連動して自動的にエージェントの顔が切り替わる。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の顔が替わるわけではない。

通常の WWW ブラウザで直接、VSA タグを含んだ WWW ページを開いても全く問題はない。VSA タグは、未知のタグとして単に無視されるだけである。この場合、ページエージェント設定のない普通の WWW ページと同じに扱われる。

3.3 VSA 属性ファイル

VSA 属性ファイルは拡張子 “.vsa” を持つテキストファイルで、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 対応テーブル
- 顔形状データ

3.4 実装

ページエージェント機能は、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 WWW 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 (一部、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。

顔画像の実時間生成、連続音声認識、規則音声合成などの基本機能はすべてクライアント側に存在し、顔画像、個人属性、および音声対話用データ等を WWW サーバから取得する。

VSA Proxy プロセスは、アクセスした WWW 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 タグを検出すると、VSA 属性ファイルの URL を顔画像生成プロセスに転送する。同プロセスは、Netscape のページアクセスと並列に WWW サーバから VSA 属性ファイルおよび JPEG

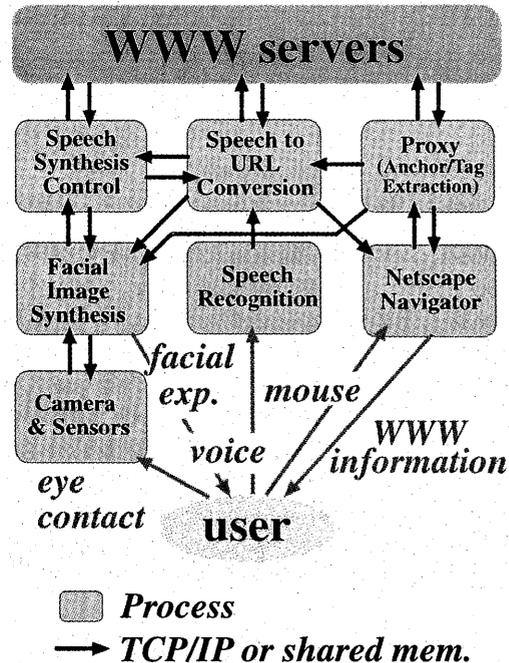


図 2: VSA implementation

圧縮された顔画像ファイルをダウンロードする。

プロトタイプでは、640×480×24 ビットの顔画像を使用した。1人のページエージェントを生成するためのオーバーヘッドは、JPEG 圧縮後の顔画像ファイルと VSA 属性ファイル、合わせて約 50～60KB である。顔画像は実際には顔の部分だけがあればよいので、さらに小さくすることも可能である。また、二つのファイルの内容は顔画像生成プロセスでキャッシュされる。例えば “Back” ボタンや「前のページ」という発話により一つ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、ページ毎にエージェントの顔が自動的に切り替わるページエージェント機能を実現した。

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, pp.35-40, 1996
- [2] H. Dohi and M. Ishizuka: “Visual Software Agent: A Realistic Face-to-Face Style Interface connected with WWW/Netscape”, *Proc. IJCAI-97 Workshop on Intelligent Multimodal Systems*, pp.17-22, 1997