

J P E G圧縮顔画像を持つ擬人化エージェントとWWWの結合*

2W-3

二階堂信夫 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

音声対話においてユーザが心理的抵抗感なくコンピュータと自然なインタラクションを実現するためには、擬人化エージェントの存在は重要である。我々が開発したビジュアルソフトウェアエージェント (VSA)[1] も含め、自然なエージェントを生成するためにはテクスチャマッピングを用いる場合が多いが、この処理はシステムの負荷が重い。そこでハードウェア JPEG ボードを用いてシステムに重い負荷をかけることなく自然感の高いエージェント像を実時間で生成する研究を進めてきた[2]。予め JPEG 形式に圧縮したさまざまな顔画像を準備しておき、実行時に任意の順序でつなぎ合わせる。本研究ではこのエージェントを WWW 用クライアントソフトウェアである Mosaic と結合し、簡単な音声対話で必要な情報を得ることができるシステムを構築した。

2 エージェントシステムの構成

擬人化エージェントシステムと WWW / Mosaic との関連を図 1 に示す。エージェントシステムは、対話処理部、ユーザ認識部、エージェント生成部から構成される。対話処理部は、ユーザからの音声による質問、依頼に対する応答文の生成処理を行い、音声で答える。またエージェント生成部に口の形を決定するための情報を伝える。ユーザ認識部は、エージェントがユーザの方向を向いて対話するように、ユーザの位置をエージェント生成部に伝える。エージェント生成部は、対話処理部、ユーザ認識部からの情報をもとにして該当するエージェントの圧縮イメージをつなぎ合わせ、ハードウェア JPEG ボードで実時間伸張処理を行いディスプレイに表示する。

ハードウェアは、シリコングラフィックス社の UNIX ワークステーション Indy と JPEG 画像圧縮ボード (Cosmo Compress) を用いた。対話処理部には市販の規則音声合成装置と、不特定話者連続音声認識装置を用いている。ユーザ認識部には、超音波距離計測器を接続しユーザの位置を把握している。

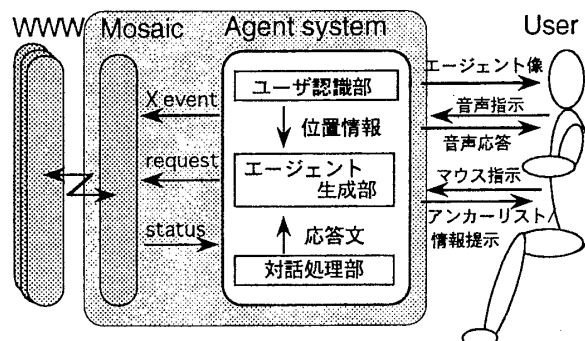


図-1 システム構成図

3 JPEGによるエージェントの生成

3.1 エージェントの生成方法

エージェントの原画像は、実際の人物の顔の画像を3次元頭部ワイヤーフレームモデルにテクスチャマップしたものを使用する。さまざまな顔の向き (1° 単位で左右 45° 上下 30°) や瞬き (2種)、口の形 (母音と 'ん' の6種) などの変形操作を加えて JPEG 形式に圧縮した。また次の方法によりデータの削減を図った。

画面分割：各顔の向きに対してフルサイズ画像 (640×480) で12種類必要とするところを、目の部分と口の部分に上下2分割しハーフサイズ画像 (640×240) 8種類で合成表示することにより、データ量を $1/3$ に削減。

データのブロック化：プロトタイプシステムでは、 1° 単位の 3×3 のブロックで圧縮イメージを共有 (実質 3° 単位) することにより、データ量を 1° 単位に比べて $1/9$ に削減した。

この結果、JPEG 圧縮と合わせて 9.2Gbyte の原画像を約 45Mbyte に削減した。このデータ量は起動時に物理メモリ上へ展開可能であり、JPEG 伸張ボードに連続的に効率良くデータを与えることができ、スムーズな顔の動き、発話に合わせた口の形の表現ができる。また、瞬きや顔の揺らぎを加えることにより、アニメによる表現や静止画では得られないより高い自然感を与えることができた。

3.2 エージェントの再生速度

ハードウェア JPEG ボードを使用すれば、 640×480 ピクセルの画像を毎秒 30 フレームの速度で伸張できる。ところが今回使用した JPEG ボードでは、上下2分割した画像を直接つなぎ合わせて表示することができない。そこで伸張した2枚の画像データ (目の部分と

*An Anthropomorphic Interface Agent generated by JPEG Compression Technique and WWW/Mosaic

†Nobuo NIKAI, Hiroshi DOHI, Mitsuru ISHIZUKA

‡Faculty of Eng., University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN

口の部分)を一度メモリ上でつなぎ合わせ、それからフレームバッファに転送している。今回のシステムでは、640×480ピクセルの大きさで毎秒9~10フレームの再生速度を得ることができた。ソフトウェアのみによるJPEG伸張、JPEGボードを用いたハードウェア伸張(圧縮データ/伸張データの転送にリングバッファを用いない場合と用いた場合)の再生速度を比較した。表1はフレームサイズ640×480ピクセル100枚の再生速度の比較結果である。経過時間は、再生開始時間と終了時間をtime()関数で、システムタイムとユーザタイムについてはtimes()関数を用いて計測した。

表-1 エージェントの再生速度比較

Decompression method		Elapsed Time(sec)	System Time	User Time
Software	Use Buffering	78.4	106.5	12.0
	No Use Buffering	17.1	433.6	9.0
Hardware	Use Buffering	10.5	100.7	13.0

(640×480pixel)×100Frame

4 エージェントとMosaicの結合

4.1 Mosaicとのインタフェース

Mosaicでは、ハイライト表示の文字列あるいはイメージ(以下アンカー)をマウスでクリックすることにより、そのアンカーにリンクされた情報を呼び出すことができる。本システムでは、ユーザとMosaicとの中間に位置するエージェントがユーザとの自然なコミュニケーションをとり、音声対話によるユーザの要求をMosaicに伝える。エージェントとMosaicとは独立したプロセスとして設計されており、プロセス間通信により必要なデータを交換する。これは、できるだけ元のMosaicの機能に影響を与えないようにするためである。マウス操作と音声操作は同等に扱われ、両者を混在させることも可能である。認識する音声コマンドは大きく次の三つに分けられる。

- キー文字列によるセレクション

音声認識でのキー文字列とMosaicのアンカー文字列との対応表を内部に持つ。ユーザの発話から抽出したキーを用いて音声によるアンカーのクリックを行う。

- インデックス番号によるセレクション

ユーザが検索するすべてのアンカーを事前に登録しておくことは不可能なため、新しいページを開くたび自動的にアンカーリスト(そのページに含まれるアンカーの一覧表)をユーザに提示し、その番号を用いて音声によるアンカーのクリックを行う。画像のような文字列を含まないアンカーにも有効である。

- 予約語によるページ制御

アンカーを選択するのではなく、予め決められた方法でページ間、ページ内の移動などの制御を行う。

4.2 実行例

本システムでの実行例を示す。

User:「石塚研究室のホームページを見せてください。」

Agent:「はい、お待ち下さい。」

(指定されたサーバと接続し、アンカーリストを表示する。)

Agent:「はい、15のアンカーがあります。」

User:「7番を見せて下さい。」

Agent:「(二階堂のホームページをオープンし、アンカーリストを表示する。)

User:「自己??を見せて下さい。」

Agent:「もう一度、お願いします。」

(音声認識に失敗した。)

User:「自己紹介を見せて下さい。」

Agent:「(自己紹介のページをオープンし、アンカーリストを表示する。)

...

5 おわりに

本研究では、ハードウェアJPEG圧縮ボードを用いて生成した自然感の高い擬人化エージェントとWWWクライアントソフトウェアの一つであるMosaicとを結合した。これによりユーザは、擬人化エージェントとの簡単な音声対話によりWWWサーバから必要な情報を引き出すことができる。計算機に対して不慣れな人や、身体的ハンデがありマウス操作が困難な人にも有効である。今回のエージェントはユーザとコミュニケーションするインタフェースエージェント的役割が主となっているが、今後利用者の問題解決や学習機能を備えた自律的なエージェントを開発する予定である。

6 参考文献

- [1] H.Dohi and M.Ishizuka: "Realtime Synthesis of a Realistic Anthropomorphous Agent toward Advanced Human-Computer Interaction", HCI'93, 1993
- [2] 二階堂・土肥・石塚: "JPEG 画像圧縮技術を用いた自然感の高い実時間擬人化エージェント像の生成", 第51回情処全 1995
- [3] 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D-II(掲載予定), 1996