グラフィクスとCAD 90－2
（1998. 4. 24）

仮想人間エージェントによるWWW上でのプレゼンテーション

乃万 司　大石京治　白鳥良男
九州工業大学 情報工学部 知能情報工学科

我々は，注記付きスピーチテキストの形の入力から，仮想人間エージェントによるプレゼンテーションを，音声出力に同期したリアルタイム3次元アニメーションとして生成するシステムを開発している．本稿では，この仮想人間エージェントによるプレゼンテーションシステムのWWW上への拡張を紹介する．WWW化のため，具体的には，(1) 遠隔画像ファイルのHTTPによる転送と仮想visual aidへの利用を可能にし，(2) 画像上の位置と名前との対応付けを行ない，(3) 入力スピーチテキストの形式を変更した．

Presentations by a Virtual Human Agent on the WWW

Tsukasa Noma　Kyoji Oishi　Yoshio Shiratori
Department of Artificial Intelligence
Kyushu Institute of Technology

We are developing a presentation system where a virtual human agent makes presentations in real time 3D animation synchronized with his speech outputs from the input of "annotated" speech texts. This article presents its extension to a WWW application including (1) transfer of remote image files via HTTP and their use for virtual visual aids, (2) definition of a mapping between positions on a image and their names, and (3) extension of input speech texts.

1 はじめに

近年，ユーザインタフェースに仮想生物エージェントを用いる試みが種々なされている．これらのエージェントの主たる役割は，情報をユーザに効果的に伝達することであり，これは広義の情報のプレゼンテーションとみなすことができる．プレゼンテーションを陽に目的としたエージェントの実施例としては，3 次元アニメーションによるプレゼンテーション [8] があるが，これはプレゼンテーションの「振付け」を手作業で行なったものである．

そこで我々は，プレゼンテーションの自動作成を目指して，コマンドを埋め込んだスピーチテキストから，仮想人間エージェントによるプレゼンテーションを，音声出力に同期したリアルタイム 3 次元アニメーションとして生成するシステムを開発している [6][7]．このシステムでエージェントは，黒板やホワイトボードなどの visual aid を仮想化した仮想ボードを用いてプレゼンテーションを行なう．この仮想ボードには，任意の画像をテクスチャとして表示できる．

我々は既に，(1) スタンドアロンの環境において，コマンド埋め込みスピーチテキストのファイルを入力とするプレゼンテーションシステム，および (2) TCP/IP ソケットを介して，クライアントからコマンド埋め込みスピーチテキストが送られると動作する一種の「インタフェースエージェントサーバ」として，本システムを実現している [6][7]．これらの機能に加え，さらに，本システムが広域ネットワーク上のデータ資源を自由に利用できるようになれば，世界中で一種の「番組発信／受信」を容易に行なうことができるようになるであろう．

そこで本稿では，その一方法として，本システムの World Wide Web (WWW) 上への拡張を紹介する．本稿ではまず，次節で従来のシステムについて概説した後，本システムを WWW に対応したアプリケーションへ拡張するための方針と実現方法を論じ，最後に結果を紹介する．

```
\board{berthapanel}
\point_idxf{berthapanel.board.bertha}
Hurricane Bertha is now to the east of
\point_back{berthapanel.board.florida}
Florida peninsula.
\point_idxf{berthapanel.board.newyork
berthapanel.board.philadelphia} New
York and Philadelphia may be hit
directly.  \gest_warning Take care.
```

図 1: 入力テキストの例

2 仮想人間エージェントによるプレゼンテーションシステム

本節では，従来から我々が開発してきた，仮想人間エージェントによるプレゼンテーションシステムの概要を紹介する．

2.1 入力

エージェントに対して何らかの指示をするとき，その指示方法は，人間にとってなるべく直観的であることが望ましい．したがって，エージェントにプレゼンテーション内容を指示する場合は，人間にとって自然なプレゼンテーション内容記述を用いることが必要であろう．我々人間がプレゼンテーションを準備する際のことを考えると，我々は，スピーチの原稿を用意し，さらにスライドを指示するなどの動作を原稿中のその場所に書き込むことが多い．

そこでシステムへの入力は，スピーチテキストにバックスラッシュで始まるコマンドを埋め込んだテキストとした (図 1)．コマンドは主にエージェントの動作に関するものであり，動作は原則としてその動作のコマンドの次の語の発話と同時になされる．現在利用可能なコマンドを表 1 に示す．コマンドの引数は {} で囲まれる．\board{} コマンドの引数は，表示される仮想ボードを表す．指示コマンドの引数は指示すべき場所を表し，複数の場所が指定された場合，それらは順に指示される．

コマンド	意味
\board{}	仮想ボードの指定
\point_down{}	手の平を下に向け指示
\point_idxf{}	人さし指で指示
\point_back{}	手の平を仮想ボードに向け指示
\point_move{}	手の平向きに動かしながら指示
\gest_givetake	提案のジェスチャ
\gest_reject	拒否のジェスチャ
\gest_warning	警告のジェスチャ

表 1: 利用可能なコマンド

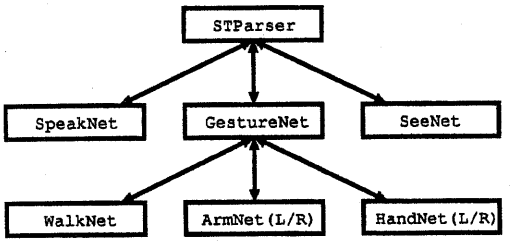

図 2: PaT-Net の構造

2.2 PaT-Net による制御

エージェントの動作は，PaT-Net (Parallel Transition Network)[2] と呼ばれる並列に動作する有限状態マシンで制御される．この有限状態マシンは，単独でも簡単な並列動作を扱え，Net 間で互いにメッセージ通信ができるように拡張されている．現在は，図 2 に示すように，9 個の PaT-Net を用いて 1 人のエージェントを制御している．矢印は Net 間のメッセージ通信を表す．

図 2 の下位の Net は人体の各部を制御する．例えば，ArmNet に指示する場所を与えると，inverse kinematics により腕が自動的にその場所を指示する．また，ArmNet に free メッセージを送ると，腕は自動的にもとの姿勢に戻る．上位の Net から指示する場所やとるべき腕の姿勢が再び指示されると，直ちにその姿勢に向かって動作しはじめる．このような一種の知的人体部品により，ある動作から次の動作への遷移を容易に扱うことが可能になる．

SpeakNet は，エージェントの口を動かすほか，発話すべきテキストを TTS (Text-To-Speech) システムに送る．最上位の STParser は，入力テキストのパーサーとして働き，他の Net を駆動させる．このようにして本システムでは，入力テキストの処理から個々の関節の動作に至るまで，PaT-Net のネットワークにより制御されている．

PaT-Net による制御，とりわけ STParser の動作については，[7] を参照されたい．

2.3 プレゼンテーションスキルと歩行動作

効果的なプレゼンテーション，すなわち情報伝達を行なうためには，エージェント側で十分なプレゼンテーションスキルを備えている必要がある．そこで本研究では，手足の位置，肩の方向，視線の向き等に関するスキルを，多くのプレゼンテーション関連文献から抽出し，エージェントの動作に組み込んだ．エージェントに組み込んだ具体的なスキルの内容とその出典については，[7] を参照されたい．

またエージェントは，同じ位置から仮想ボード上のすべての場所を指示できるわけではない．そこでエージェントは，次の指示コマンドを先読みし，指示に使う手 (右／左) と事前の移動の必要性を決定する．指示する手は，現在の体の位置からの移動量や仮想ボードの視点からの隠蔽具合を考慮したヒューリスティックスによって決定される．

従来の歩行動作生成に関する研究は，直線上や滑らかな曲線上の歩行を対象とするものがほとんどであった．しかし，プレゼンテーション時の人間の歩行は，数歩の内に，前／横／後ろ向きの歩行や方向転換などが組み合わされたものである．このような種々の移動方法を扱える locomotion engine として VRLOCO[4] があるが，これは体の位置と向きの時系列データの入力から locomotion を生成するものであり，本

システムで要求されているように移動先が指定されただけで locomotion を生成するものではない．そこで本システムでは，移動先を指定するだけで，これらの種々の歩行動作を統合した形で locomotion を生成する locomotion engine を WalkNet (図 2) として実現した．本システムのエージェントは，現在の体の位置から次に指示すべき場所に手が届かない場合は，WalkNet を用いて，事前に体を移動させる．

2.4 実現

本システムの実現には，SGI 製 Onyx/Reality 上の *Jack*[1] システムを用い，音声出力用 TTS には Entropic Research 製の TrueTalk を用いた．なお，TrueTalk の TTS サーバは，別のマシン (SGI 製 Indigo2) で実行させた[1]．その結果，プレゼンテーションのアニメーションが音声と同期してリアルタイムに (毎秒 30 フレーム) 生成された．

第 1 節で述べたように，当初採られたシステムの実現形態は，以下の二種類であった [6][7].

(1) **スタンドアロン環境のプレゼンテーションシステム**
コマンド埋め込みスピーチテキストのファイルを入力としてプレゼンテーションを行なう．

(2) **TCP/IP ソケットを介したインタフェースエージェントサーバ**
クライアントからコマンド埋め込みスピーチテキストが送られると，本システムがインタフェースエージェントを動作させる．

以下本稿では，本システムの第三の実現形態として，WWW 上のアプリケーション化を試みる．

[1]これは，計算負荷の軽減のためではなく，使用していた Onyx では TrueTalk による音声出力が不可能だったためである．

3 システムの WWW 化

本システムを WWW 化するためには，次の機能が実現されなくてはならない．

(1) WWW のブラウザ上でシステムを起動できること．

(2) 入力スピーチテキストファイルを遠隔システムから転送できること．

(3) 仮想ボード用の画像ファイルを遠隔システムから転送できること．

(4) 仮想ボード上の位置と名前とを対応付けられること．

それぞれの実現方法を以下で述べる．

3.1 システムの起動と入力ファイルの転送

(1) および (2) については，本システムを，ブラウザのプラグインではなく，ブラウザから起動される外部アプリケーションとして利用するものとすれば，WWW サーバ側でメディアタイプを登録し，ブラウザ側でそのメディアタイプに対し本システムが起動されるよう登録すれば十分である[2]．

3.2 画像ファイルの転送

(3) については，\board{} コマンドの引数で画像ファイルの URL を指定させ，その URL のファイルを (HTTP で) 転送させることにした [9]．転送された画像ファイルは，John Cristy の ImageMagick を用いて，テクスチャ用のピクセル配列に変換される．なお現在は，キャッシング等は行なっていない．

[2]現在の実験では，入力スピーチテキストのメディアタイプを application/v-presenter，ファイル名拡張子を .vps と定めている．

```
bertha:0.70:0.79
florida:0.44:0.72
newyork:0.56:0.11
philadelphia:0.54:0.16
```

図 3: VBM ファイルの例

3.3 仮想ボード上の位置と名前の対応

以前のシステムでも，仮想ボード上の位置を名前で参照することができていた．これは，仮想ボードを *Jack* システムの figure (図形) として定義し，その定義中で位置と名前の対応を与えていたためである．しかし，遠隔画像ファイルを直接参照する場合，それらの画像ファイルに，そのような位置と名前の対応情報を期待することは出来ない．

そこで，図 3 のように，画像上の正規化座標と名前との対応を与える VBM (Virtual Board Mapping) ファイルを定め，このファイルを用いて，仮想ボード上の位置と名前の対応を与えることにした．例えば図 3 では，`bertha` とは，画像の左上を $(0,0)$，右下を $(1,1)$ とする座標系で，$(0.70, 0.79)$ の位置を表している．使用する VBM ファイル名は，`\board{}` コマンドの引数中で，画像ファイルの URL の後に与えることにした．

3.4 入力ファイルの変更

上記の変更に伴い，入力のコマンド埋め込みスピーチテキストも変更される．変更後の入力テキストの例を図 4 に示す．ここでは，図 3 の VBM ファイルが `\board{}` コマンドの引数中で `ecoast.vbm` として参照されている．この例では，VBM ファイルはローカルなファイルだが，URL を指定して VBM ファイルに遠隔ファイルを用いることも出来る．

3.5 表示システムの変更

2.4 節で述べたように，元のシステムでは，人体モデルの定義と表示に *Jack* システムを用いていた．しかし，*Jack* システムは，もともと対話型の動作制御を目的として設計されており，本システムのようなリアルタイムアニメーション向きではない．そこで，*Jack* システムのモデル定義をほぼそのまま採用したより「軽い」モデル定義／表示システムを OpenGL を用いて実現し，今回のシステムで採用した．

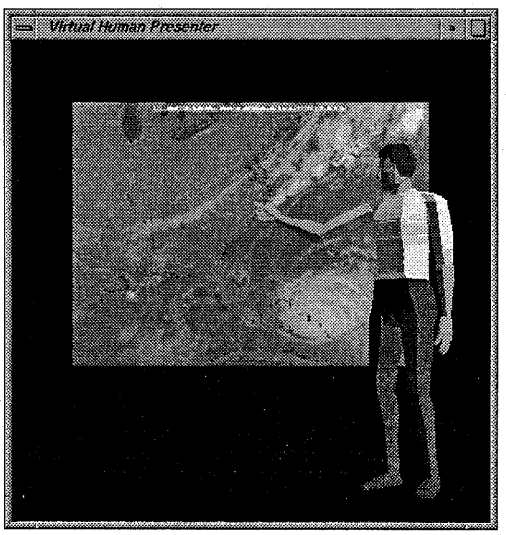

図 5: プレゼンテーションの例

4 結果

上記システムを実現し，WWW 上で，仮想人間エージェントによるプレゼンテーションが可能であることを確認した．その表示例を図 5 に示す．現在は，従来とほぼ同じく，SGI 製 Onyx2/Reality と (別マシン (SGI 製 O2) で TTS サーバを実行させている) Entropic Research 製 TrueTalk との組合せで実験を行なっている．

しかし，表示システムが変更されたこともあり，SGI 製 O2 等の安価なグラフィックスワークステーション上で，しかも，TTS サーバを同一マシンで実行させた場合でも，実用的な表示速度が得られる見通しが得られている．

```
\board{http://www.ncdc.noaa.gov/pub/data/images/hurricane-bertha-avhrr-vis.gif,
ecoast.vbm} \point_idxf{bertha} Hurricane Bertha is now to the east of
\point_back{florida} Florida peninsula. \point_idxf{newyork, philadelphia}
New York and Philadelphia may be hit directly. \gest_warning Take care.
```

図 4: 変更後の入力テキストの例

5 むすび

本システムにより，一種の注釈付きスピーチテキストを WWW サーバに置くだけで，世界中に「番組発信」が出来，また，ブラウザ上でそのテキストへのリンクをクリックするだけで，どこでもその番組を受信することが出来るようになった．

本システムの注釈付きスピーチテキストによるプレゼンテーション生成は，MPEG や AVI などの動画ファイルを直接送る場合に比べ，(1) 制作が容易で，(2) データ転送量が少なく，(3) 視聴者側でその嗜好に合わせてエージェントの種類や動作を変えられるという特長がある．特に最後の特長から，本システムは，Negroponte がその著書 *Being Digital* [5] の中で述べた「送られてきたビット情報を local computing intelligence を用いて視聴者の趣味に合わせた TV 番組に変える」技術への第一歩になりえると考えている．

また，本システムは，3 次元アニメーションを用いており，[3] などの 2 次元画像に基づくインタフェースエージェントに比べ，動作の柔軟性と拡張性においてはるかに優れているものと考えられる．

謝辞

本研究のプログラムの一部は，高田純司君 (九州工業大学情報工学部知能情報工学科平成 9 年度卒業) の卒業研究として作成されたものです．また，本研究は第一著者と University of Pennsylvania の Norman I. Badler 教授との共同研究を発展させたものであり，第一著者の University of Pennsylvania での研究は，文部省在外研究員として行なわれたものです．この共同研究には，米国政府諸機関およびジャストシステム (株) からご援助いただきました．ハリケーン Bertha の衛星写真は，NOAA / National Climatic Data Center によるものです．

参考文献

[1] Badler, N.I., Phillips, C.B., and Webber, B.L., *Simulating Humans: Computer Graphics Animation and Control*, Oxford University Press, 1993.

[2] Cassell, J. et al., "Animated Conversation: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents," *Proc. SIGGRAPH 94*, pp. 413–420, July 1994.

[3] Gibbs, S., Breiteneder, C., de Mey, V., and Papathomas, M., "Video Widgets and Video Actors," *Proc. UIST 93*, pp. 179–185, 1993.

[4] Ko, H. and Cremer, J., "VRLOCO: Real-Time Human Locomotion from Positional Input Streams," *Presence*, Vol. 5, No. 4, pp. 367–380, 1996.

[5] Negroponte, N., *Being Digital*, Random House, 1995.

[6] Noma T. and Badler, N.I., "A Virtual Human Presenter," *Proc. IJCAI-97 Workshop on Animated Interface Agents*, pp. 45–51, Aug. 1997.

[7] 乃万司, Badler, N.I., "プレゼンテーションエージェントのスキルと駆動," 第 3 回知能情報メディアシンポジウム論文集, pp. 231–238, Dec. 1997.

[8] Thalmann, N.M. and Kalra, P., "The Simulation of a Virtual TV Presenter," In: *Computer Graphics and Applications (Proc. Pacific Graphics '95)*, pp. 9–21, World Scientific, 1995.

[9] Wong, C. (法林浩之監訳, 須田隆久訳), Web クライアントプログラミング, オライリー・ジャパン, 1997.